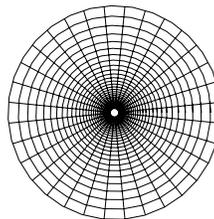




UNIVERSIDADE TÉCNICA DE LISBOA
INSTITUTO SUPERIOR TÉCNICO



SEGUIMENTO BINOCULAR DE ALVOS MÓVEIS
BASEADO EM IMAGENS LOG-POLAR

ALEXANDRE JOSÉ MALHEIRO BERNARDINO

(Licenciado)

Dissertação para obtenção do Grau de Mestre em
Engenharia Electrotécnica e de Computadores

Orientador Científico:

Doutor José Alberto Rosado dos Santos Victor

Constituição do Júri:

Doutor João José dos Santos Sentieiro

Doutor Helder de Jesus Araújo

Doutor José Alberto Rosado dos Santos Victor

Lisboa, Dezembro de 1996

Resumo

Muitas tarefas em robótica podem beneficiar perceptualmente do seguimento de alvos móveis no espaço visual. Esta tese tem como objectivo o estudo e projecto de um sistema de seguimento de alvos móveis genéricos. Apresenta-se a implementação do sistema projectado numa cabeça robótica de visão activa.

Generalidade, fiabilidade e funcionalidade em tempo-real são preocupações principais no projecto do sistema. Recorre-se à utilização de informação visual de fácil e rápida extracção, sem proceder ao reconhecimento do alvo. São integrados aspectos motivados pelo estudo de sistemas visuais biológicos: a geometria binocular, a representação de imagens de resolução não uniforme e a decomposição em movimentos oculares separáveis.

É utilizada a transformação log-polar na representação das imagens, que proporciona elevada resolução no centro e menor resolução na periferia. Esta representação revela-se mais adequada aos comportamentos de seguimento do que a habitual representação cartesiana. A decomposição dos movimentos oculares em vergência e seguimento simplifica o projecto do sistema e a aquisição da informação visual. A vergência é atingida através de informação *stereo* proporcionada pela geometria binocular do sistema e possibilita a seguimentação do alvo. Estimativas da posição e velocidade do alvo são utilizadas no controlo dos movimentos de seguimento.

Palavras Chave: visão robótica, percepção, sistemas visuais biológicos, representação log-polar, vergência e seguimento, controlo em tempo-real.

Abstract

Tracking moving targets in the visual field can improve perception in many robotic tasks. This Thesis addresses the design of a tracking system for generic visual targets. We present an integrated system using a binocular active vision head.

Generality, reliability and real-time performance are the main concerns of the design. The control process avoids target recognition by using low-level visual cues, which are fast and easy to compute. The design is based on several aspects of biologic visual systems: binocular geometry, space variant imaging and motion decomposition into separable eye movements.

Images are represented using the log-polar mapping, which provides high resolution in the center of the visual field and lower resolution in the periphery. These images have advantages over the usual cartesian representation in many aspects of the tracking problem. The design and the acquisition of visual information are simplified by using separable eye movements: vergence and pursuit. Vergence is attained by binocular information and enables the segmentation of the target. Target retinal position and velocity are used to control the pursuit movements.

Keywords: robot vision, perception, biologic visual systems, log-polar imaging, vergence and pursuit, real-time control.

Agradecimentos

Em primeiro lugar quero agradecer profundamente ao Prof. José Alberto Santos Victor pela orientação decisiva do trabalho aqui apresentado. Pela disponibilidade constante e pelo apoio a vários níveis que sempre me proporcionou. Por me ter mostrado que organização e perseverança são dois aspectos fundamentais na abordagem de problemas complexos. E ainda por revelar sempre bom senso e honestidade na discussão científica.

A todos os colegas do grupo de visão do I.S.R. quero agradecer o espírito de grupo criado e a disponibilidade para longas trocas de ideias. Em particular: ao Gaspar por me ter familiarizado com o *software* e *hardware* do laboratório de visão, ao César pela boa disposição e criatividade contagiantes, ao António pela ajuda na resolução de muitos problemas, ao Nuno Pereira pelas melhorias introduzidas na Medusa, ao Vitor pela companhia nos lanches da sala de convívio, ao Carlos pela companhia fora de horas e ao Etienne pelo seu espírito matemático.

A todos os outros colegas do ISR agradeço as conversas de corredor e ajudas pontuais, e que por serem tantos não posso enumerar.

À Marta, pelo lindo sorriso que me deixou mapear para log-polar.

Ao Marco, pela amizade que nos tem unido ao longo de tantos anos.

Ao Instituto de Sistemas e Robótica pelas excelentes condições de trabalho que me proporcionou.

À Junta Nacional de Investigação Científica, pelo financiamento parcial desta tese, através

da bolsa *PRAXIS XXI / BM / 1673 / 94*.

De uma forma muito especial quero exprimir os meus agradecimentos a todas as pessoas que me apoiaram moralmente durante este trabalho. Aos meus Pais, pelo amor e incentivo que me deram. Ao meu irmão, pela companhia e amizade sempre presentes. Às minhas avós, pela tolerância com que aceitaram as minhas ausências.

Finalmente, à Helena. Pela compreensão e encorajamento. Pela companhia e motivação. Pelo amor dado ao longo destes anos.

Índice

1	Introdução	1
1.1	Motivação	2
1.1.1	Constituição da retina	2
1.1.2	Movimentos oculares	3
1.1.3	Processos visuais básicos	4
1.1.4	Mecanismos de atenção	5
1.1.5	Cognição e percepção	6
1.2	Objectivos	7
1.3	Trabalho relacionado	9
1.4	Contribuições originais	11
1.5	Organização da tese	12
2	Um observador artificial	15
2.1	A cabeça robótica Medusa	15
2.2	Movimentos oculares	16
2.2.1	Movimentos oculares e estímulos visuais na fixação	17
2.3	Medusa: aspectos geométricos	18
2.3.1	Geometria binocular	20
2.3.2	Utilização de coordenadas esféricas	23
2.4	Formação de imagem e disparidade binocular	25
2.4.1	Modelo de formação de imagens	26
2.4.2	Disparidade	27
2.4.3	Horóptero geométrico	29

3	Retina não uniforme.	33
3.1	Imagens com resolução não uniforme.	33
3.1.1	Resolução	33
3.1.2	Resolução de imagens contínuas.	34
3.1.3	Resolução de imagens discretas.	39
3.1.4	Aquisição de imagens discretas com resolução não uniforme	42
3.1.5	Imagens reamostradas.	45
3.2	Transformação log-polar	46
3.3	Emulação da retina	51
3.3.1	Parâmetros da transformação	52
4	Vergência	59
4.1	Estímulos para controlo de vergência	60
4.2	Correlação para controlo de vergência	61
4.2.1	Medidas de correlação entre imagens	62
4.2.2	Correlação em imagens de resolução não uniforme	63
4.3	Algoritmos de controlo de vergência	67
4.3.1	Controlo por optimização local	68
4.3.2	Controlo por optimização global	70
4.4	Experiências e resultados	74
5	Seguimento	79
5.1	Hipótese de vergência permanente	79
5.1.1	Representação ciclópica	81
5.2	Estímulos de seguimento	82
5.3	Controlo de seguimento	82
5.3.1	Cinemática e sensoriamiento	84
5.4	Estimação de movimento	87
5.4.1	Erro de posição	88
5.4.2	Escorregamento retinal	90
5.4.3	Projecto do controlador de seguimento	94

6	Integração e coordenação	97
6.1	Segmentação do alvo	97
6.1.1	Fusão binocular	99
6.1.2	Segmentação por disparidade nula	99
6.1.3	Influência da representação log-polar	102
6.1.4	Algoritmo de segmentação	105
6.2	Vergência e seguimento	106
7	Resultados	111
7.1	Experiências em cadeia aberta	112
7.2	Experiências com movimento conhecido	114
7.3	Experiências com movimento desconhecido	118
8	Conclusões	129
8.1	Trabalho futuro	131
A	Transformação log-polar	133
B	Dimensionamento dos controladores de seguimento	135
B.1	Controlo de posição	136
B.2	Controlo de velocidade	136

1. Introdução

A visão é um sentido extremamente potente e versátil, em grande parte devido à riqueza contida na informação que processa. Cor, forma, textura e luminosidade são propriedades da informação visual que se relacionam de uma forma muito directa com os elementos constituintes do meio envolvente. A utilização deste tipo de informação pode trazer importantes benefícios em tarefas que requerem interacção com o ambiente, particularmente em robótica.

A utilização de sistemas de visão robóticos em meios industriais ou laboratoriais tem sido, em muitos casos, bem sucedida. A implementação de tais sistemas é facilitada pelo facto destes ambientes serem suficientemente estruturados e previsíveis. No entanto, em meios desconhecidos ou dinâmicos, onde é necessária uma constante adaptação às condições ambientais e reacções rápidas a acontecimentos imprevistos, a aplicação destes sistemas não tem tido os resultados desejados. Falta de robustez, inflexibilidade e lentidão de resposta são algumas das deficiências apresentadas, de um modo geral, por sistemas artificiais que operam em ambientes complexos [41].

Em contrapartida, encontramos no mundo biológico um grande conjunto de seres vivos que interage diariamente com tais ambientes, sendo o ser humano um exemplo de elevada capacidade e desempenho em tarefas quotidianas. Em particular no campo da percepção visual, tarefas com importantes implicações em robótica como a localização, a fixação e o seguimento de objectos, são consideradas triviais devido à simplicidade e fiabilidade com que o ser humano as executa, mas de difícil implementação em sistemas artificiais devido ao dinamismo e diversidade de situações possíveis.

Um agente artificial capaz de efectuar tarefas perceptuais deste tipo e de se integrar num sistema robótico mais complexo, seria de grande utilidade em ambientes dinâmicos

e/ou desconhecidos. Nesta tese, um agente desse tipo tem a designação genérica de **observador artificial**, e o seu projecto é motivado em grande parte por características de sistemas visuais biológicos.

1.1 Motivação

O desempenho do sistema visual humano depende de uma grande diversidade de factores que vão desde a constituição física dos olhos aos aspectos cognitivos e motivacionais, passando pela geometria binocular e pelo controlo dos movimentos oculares [21]. Naturalmente, o estudo e compreensão de um sistema tão complexo terão que ser considerados como objectivos a longo prazo. Há que analisar cada um dos factores individualmente e tentar compreender as suas interdependências de modo a facilitar a integração de soluções parciais.

1.1.1 Constituição da retina

Uma das características mais importantes do sistema visual humano e de outros seres vivos reside na distribuição não uniforme dos fotoreceptores na retina. A retina humana é constituída por 100 a 200 milhões de fotoreceptores¹ [68], e que se distribuem com grande densidade na zona central denominada por **fóvea**, enquanto que, em direcção à periferia, a sua quantidade por unidade de área diminui significativamente. Existem cerca de 300 vezes menos fotoreceptores na periferia do campo visual do que na fóvea [25].

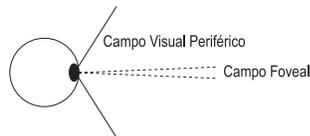


Fig. 1.1: A zona do campo visual de elevada resolução, correspondente à fóvea, tem uma amplitude muito pequena (cerca de 2 graus) relativamente ao campo visual periférico (cerca de 140 graus) [25].

Ao permitir obter elevadas resoluções na zona central da imagem observada e um

¹elementos que convertem a informação luminosa em sinais neuronais transmitidos ao cérebro.

campo visual bastante vasto, esta distribuição dos fotoreceptores estabelece duas competências importantes para a retina: a fóvea ocupa-se da inspecção cuidada da zona observada e a periferia efectua a selecção de outros ponto de interesse no restante campo visual.

A solução natural encontrada para a constituição da retina, resulta de um compromisso entre a quantidade limitada de recursos existentes (número de fotoreceptores e capacidade de processamento), e os requisitos de resolução e de amplitude de campo necessários a um bom desempenho visual. Para um mesmo número de elementos sensitivos, a utilização de uma distribuição uniforme com a mesma resolução da existente na fóvea, produziria um campo visual muito reduzido, ou então, para se ter um idêntico campo visual, seria necessária uma quantidade extremamente elevada de fotoreceptores. Neste último caso, só um cérebro humano com cerca de 15 toneladas teria a capacidade de processamento necessária para lidar com tanta informação [2]!

1.1.2 Movimentos oculares

Uma das principais razões da existência de movimentos oculares em muitos sistemas visuais biológicos é, precisamente, a densidade variante no espaço dos fotoreceptores na retina. Como apenas o centro da área observada tem elevada resolução, a inspecção cuidada de outras zonas de interesse no campo visual requer o redireccionamento da fóvea para essas zonas através de movimentos oculares. Combinando a existência de uma pequena zona de alta resolução na retina com mecanismos de selecção de zonas de interesse na visão periférica e a capacidade de movimentar os olhos, torna-se possível obter informação de alta resolução sobre uma extensa área do campo visual.

Os movimentos oculares mais influentes em sistemas biológicos são os movimentos sacádicos, os movimentos de vergência e os movimentos de seguimento lento [53]. Cada movimento tem uma competência específica na percepção visual, o que revela uma estrutura modular na constituição do sistema oculomotor. Os movimentos sacádicos alteram rapidamente a direcção de observação, de modo a obter informação visual em pontos específicos do espaço. Os movimentos de vergência permitem manter a fixação em objectos que se desloquem em profundidade e os movimentos de seguimento lento compensam o

deslocamento do alvo noutras direcções. Por si só, a utilidade destes movimentos é reduzida mas, em conjunto cooperam na execução de muitas tarefas visuais importantes, em particular na inspecção, detecção e seguimento de objectos no espaço visual.

1.1.3 Processos visuais básicos

Os insectos têm sido, para alguns autores [11, 4, 55], modelos de comportamento para aplicações em robótica. Os sistemas visuais destes animais são relativamente simples mas, mesmo assim, conseguem ser eficazes em ambientes dinâmicos e não estruturados, que são ambientes tipicamente difíceis para agentes artificiais. O seu desempenho perceptual depende de um conjunto reduzido de processos visuais simples, com base nos quais se atingem comportamentos mais complexos. Em [38] são referidos alguns processos visuais básicos comuns entre seres humanos e certos tipos de insectos, como sejam: a detecção de alvos de interesse, a fixação de alvos estáticos, o seguimento de alvos móveis e a deambulação visual. Qualquer destes processos é composto por combinações de movimentos oculares.

Os processos de fixação e de seguimento têm como objectivo manter os objectos de interesse dentro das zonas de maior acuidade visual na retina, compensando quer os movimentos do alvo, quer do observador. A distinção entre os dois processos advém das características de movimento do alvo. Considera-se fixação no caso em que o alvo permanece estático, sendo o processo de fixação responsável pela compensação do movimento próprio do observador. Por outro lado, o processo de seguimento efectua a compensação do movimento do alvo. No decorrer desta tese consideramos estes dois processos como idênticos embora em termos biológicos isso não seja correcto — a fixação pode utilizar informação sobre o movimento próprio obtida por outros meios sensoriais que não a visão, enquanto que o seguimento utiliza principalmente informação visual. Dado que se pretende usar apenas sensoriameto visual, a utilização dos termos **fixação** e **seguimento** é usada indistintamente para designar o mesmo processo: a manutenção do alvo no centro das retinas.

Os processos de detecção de alvos e de deambulação visual estão inerentes a comportamentos de exploração ou de procura. São compostos por alterações rápidas da direcção

de observação, intervaladas com períodos de estabilização visual para permitir recolha de informação. O que distingue os dois processos é o objectivo prioritário de cada um. Enquanto na detecção de alvos existe conhecimento prévio do alvo a detectar, na tarefa de deambulação visual isso não acontece.

1.1.4 Mecanismos de atenção

A grande variedade de estímulos visuais com que somos confrontados no dia a dia e o número limitado de recursos visuais e computacionais de que dispomos, requer a análise selectiva dos objectos presentes no campo visual. A **atenção visual** é a capacidade de detectar rapidamente partes interessantes da imagem retinal de modo a reduzir a quantidade de informação a processar [42]. Em sistemas visuais com fóvea, a detecção de um objecto de interesse geralmente conduz a um redireccionamento da fóvea para o objecto detectado.

Os mecanismos de atenção visual envolvem uma grande componente cognitiva e motivacional. A experiência visual do observador, a capacidade de raciocínio e o seu estado motivacional influenciam em grande parte as suas reacções visuais perante o ambiente. Esta é a principal dificuldade com que nos deparamos na análise dos mecanismos de atenção.

Também noutras tarefas visuais, o conhecimento adquirido e a capacidade de raciocínio constituem factores muito importantes no desempenho da visão. Somos capazes de reconhecer as dimensões de uma cena através do conhecimento *a priori* da dimensão de um dos objectos presentes, ou descobrir a direcção de deslocamento de um avião simplesmente pela sua pose. Actualmente, a implementação de tais tarefas em sistemas artificiais é possível à custa do desenvolvimento de rotinas específicas, mas a grande quantidade e diversidade de tarefas que envolvem uma componente cognitiva, tornaria exaustiva esta abordagem. Estima-se que os humanos sejam capazes de reconhecer cerca de 100 000 objectos em 100-200 ms [42]. Só a compreensão mais profunda dos nossos mecanismos cognitivos e a sua formalização em termos mais rigorosos, poderá permitir abordar estes problemas de forma adequada.

1.1.5 Cognição e percepção

A primeira grande distinção que podemos fazer entre tarefas visuais está relacionada com a existência, ou não, de intervenção cognitiva aparente. A exclusão da componente cognitiva reduz em grande medida a capacidade final do observador artificial mas muitas tarefas importantes em robótica não requerem capacidade cognitivas avançadas [54]. Entre essas tarefas podemos considerar a detecção de movimento, o seguimento de alvos, o cálculo do movimento próprio, a detecção de obstáculos, cálculo de profundidade, entre outras. Utilizam apenas indícios visuais de baixo nível ou, como designados em [23], indícios pré-categóricos². Exemplos de indícios deste tipo são a disparidade binocular, a desfocagem da imagem e o fluxo óptico.

Uma análise mais detalhada dos processos visuais básicos revela características importantes sobre o nível neurológico a que cada um é controlado. Por exemplo, tanto na detecção de alvos como na deambulação visual, existe uma componente voluntária que é fortemente dependente do estado motivacional do indivíduo. Já na fixação e no seguimento de alvos predominam mecanismos involuntários: o observador determina o ponto a fixar mas abstrai-se do modo como mecanismos neuronais de mais baixo nível executam a fixação.

Esta análise permite concluir a existência de dois grandes grupos de mecanismos responsáveis pelo comportamento do sistema oculomotor. De um lado temos os **mecanismos voluntários**, desencadeados por impulsos motivacionais. São responsáveis pela alteração da atenção visual entre zonas distintas do espaço visual, através de movimentos sacádicos. De outro lado temos os **mecanismos involuntários**, responsáveis de um modo geral pela fixação da atenção em determinadas zonas do campo visual e são controlados a um nível neuronal não cognitivo.

De acordo com estas ideias, em [15] é sugerida uma decomposição dos comportamentos oculares em dois mecanismos de atenção principais: a **fixação de atenção** e a **mudança de atenção**. A coordenação entre estes mecanismos é modelizada de uma forma muito simples, de acordo com o esquema da Figura 1.2, onde as setas representam transições entre os dois comportamentos.

²do inglês, *pre-categorical cues*.

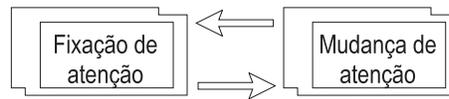


Fig. 1.2: Decomposição dos comportamentos oculares em dois mecanismos de atenção principais.

A complexidade do sistema visual devida aos aspectos motivacionais fica repartida entre o comportamento de mudança de atenção e a gestão das transições entre os comportamentos. O comportamento de fixação de atenção é constituído por mecanismos neuronais de baixo nível e responsável pelos movimentos oculares involuntários.

1.2 Objectivos

O objectivo deste trabalho, consiste no projecto de um sistema de fixação de atenção capaz de efectuar de uma forma fiável o seguimento de alvos móveis genéricos. Pretende-se que o sistema de fixação seja capaz de manter as câmaras dirigidas para o alvo, independentemente da sua forma e do seu movimento relativo.

Não é objectivo desta tese considerar o problema da selecção do ponto a fixar. Admite-se a existência de um processo de inicialização que se encarrega de colocar o objecto a fixar na direcção corrente de observação.

A utilidade de fixar e seguir alvos móveis no espaço visual revela-se em muitos aspectos. Ao manter o objecto de interesse no centro do campo visual, ou seja, na zona de maior resolução da retina, o processo de seguimento permite maximizar a quantidade de informação extraída. Ao mesmo tempo, também o deslocamento permitido para o alvo dentro do campo visual, de um instante para outro, é maximizado.

Em robótica, a utilidade da fixação visual revela-se em muitas tarefas importantes tais como a reconstrução do movimento tridimensional do objecto fixado, a estimação do movimento próprio do observador [26] ou o reconhecimento de objectos [18].

A fixação visual implementada em sistemas artificiais, também traz vantagens a outros níveis [23]. Por exemplo, quando um objecto se localiza próximo dos eixos ópticos de cada câmara, é possível utilizar hipóteses simplificativas, tal como o uso da projecção ortográfica em vez da projecção perspectiva. Por outro lado os pontos pertencentes ao objecto fixado

têm valores de disparidade baixos o que permite a utilização de algoritmos *stereo* rápidos e simples.

No projecto do sistema de seguimento, dois aspectos de sistemas visuais biológicos vão ter especial relevância: a **constituição da retina** e o **controlo oculomotor**.

À excepção de alguns casos [49, 65, 27, 60], a maioria dos sistemas de visão artificial utilizam representações de imagem constituídas por uma amostragem uniforme do espaço visual. No entanto, para sistemas com a capacidade de alterar a direcção de observação, o processamento completo das imagens com resolução uniforme pode tornar-se supérfluo: geralmente estamos interessados em obter informação em locais específicos do campo visual e não no seu todo. Uma das soluções possíveis consiste em definir janelas de processamento que são colocadas nas zonas de interesse, em cada instante, permitindo um processamento localizado da informação. A solução encontrada na biologia, contendo um foco de atenção (fóvea) direccionado para as zonas de interesse por movimentos oculares, tem a vantagem de não necessitar de representar toda a informação contida nas imagens cartesianas. A utilização de uma estratégia semelhante por parte dos sistemas artificiais poderá trazer importantes benefícios, quer ao nível da redução de informação a processar quer ao nível do aproveitamento de propriedades das topologias não uniformes no processamento de imagem [14].

O controlo oculomotor designa de uma forma genérica como os estímulos visuais são utilizados no controlo dos movimentos oculares. Em muitos aspectos a motivação biológica sugere soluções que seguem os paradigmas da Visão Activa e Visão Objectiva³ [1, 8, 9]. A utilização pistas visuais pré-categóricas de fácil e rápida extracção, específicas para o controlo de cada movimento ocular, e a ligação muito directa entre os estímulos visuais e movimentos como a vergência e o seguimento lento, conduzem a comportamentos que se adaptam facilmente a alterações no campo visual, permitindo uma elevada reactividade em situações dinâmicas e imprevistas, de grande importância em robótica [17, 6].

³em inglês, *Purposive Vision*.

1.3 Trabalho relacionado

Vários autores têm perseguido o objectivo de efectuar o seguimento de alvos móveis em sistemas de visão artificiais. Alguns trabalhos apresentam características comuns com o sistema que apresentamos nesta tese, como por exemplo: a utilização de um sistema binocular de visão activa, a generalidade do movimento e da forma admitida para os alvos de seguimento, o controlo de movimento por retroacção de informação visual e o desempenho em tempo real. Para além destes trabalhos, que se relacionam de uma forma global com o trabalho apresentado nesta tese, existem muitos outros que se relacionam com técnicas individuais aqui utilizadas, por exemplo a representação log-polar, técnicas de correlação, cálculo de fluxo óptico, controlo de sistemas dinâmicos, etc. Estes trabalhos serão referidos pontualmente no decorrer da tese.

Em [44, 43, 52] são apresentados vários aspectos do trabalho que vem sendo desenvolvido na cabeça robótica “Yorick”. Este sistema utiliza basicamente pistas monoculares tais como *matching* de cantos e o fluxo óptico. De entre as capacidades que têm vindo a ser implementadas neste sistema é de referir a localização e o seguimento de alvos móveis, assim como o redireccionamento da direcção de observação com predição da velocidade. Outra característica interessante é a discriminação entre objectos que se movem independentemente. No entanto, esta característica admite que o movimento do fundo advém apenas da rotação conhecida das câmaras, o que impede a utilização do sistema em veículos cujo movimento não é conhecido.

O trabalho apresentado em [46, 47, 48] refere-se a um sistema de seguimento de alvos baseado na cooperação de processos oculares e implementado na cabeça robótica “KTH”. A informação de profundidade do alvo é extraída com base na integração de duas pistas visuais – a disparidade e a acomodação visual – enquanto o seguimento é efectuado por correlação de zonas das imagens. A utilização de acomodação, sendo uma pista monocular, permite a manutenção da vergência mesmo quando o objecto se encontra ocluído relativamente a uma das câmaras.

Uma outra arquitectura de controlo para cabeças robóticas é descrita em [3], e inclui os graus de liberdade ópticos. À semelhança do trabalho anterior, o comportamento de vergência beneficia da integração de múltiplos indícios visuais. A utilização de técnicas de

correlação é generalizada a todos os processos visuais. No entanto não é concretizada uma implementação integrada do sistema, sendo apresentados resultados apenas em situações muito simples.

Em [62] é apresentado um sistema de seguimento bastante completo que funciona à taxa de vídeo (25Hz), baseado em técnicas de correlação a diversas escalas. No entanto, como estas técnicas falham na presença de oclusões, utilizam-se também mecanismos de segmentação do alvo baseados em fluxo óptico e disparidade. Alguns algoritmos utilizados são computacionalmente exigentes mas a utilização de *hardware* específico de processamento de imagem permite atingir elevados ritmos de amostragem.

Um dos primeiros sistemas de seguimento binocular de alvos móveis genéricos a apresentar funcionalidade em tempo-real (15Hz) é descrito em [24]. Utiliza a informação binocular de profundidade e propriedades da fusão binocular para atingir o seguimento de alvos isolados no espaço visual. A forma como está implementado assegura uma correcta integração entre os comportamentos de vergência e de seguimento, ou seja, o objecto sob vergência, será o objecto escolhido para o seguimento. Este trabalho tem diversas semelhanças com o que é apresentado nesta tese mas utiliza imagens cartesianas, o que constitui uma diferença fundamental.

O sistema descrito em [20, 19] é um dos poucos sistemas existentes que utiliza a representação log-polar para as imagens adquiridas. O controlo do seguimento é feito com base em informação obtida a partir da estimação de um modelo afim para o fluxo óptico do alvo. A componente de expansão/contractão do modelo é relacionada com o movimento do alvo em profundidade e a componente de translação corresponde ao movimento lateral. Este trabalho ilustra de uma forma concreta a utilidade da representação log-polar no cálculo do fluxo óptico e a utilização desta medida no controlo dos movimentos oculares.

Estes trabalhos têm aspectos semelhantes de um ponto de vista externo, ao abordarem o problema de seguimento segundo o paradigma da visão activa. Informação visual de baixo nível é utilizada de uma forma objectiva e directa no controlo do processo de seguimento que se deseja que seja o mais geral e robusto possível. O que os diferencia internamente é a forma como pretendem atingir estes objectivos. Tipos de estímulos sensoriais, arquitecturas de controlo, integração de medidas e representações de imagem são características internas que podem provocar diferenças significativas de desempenho.

1.4 Contribuições originais

Uma das principais contribuições desta tese relaciona-se com a aplicação da geometria log-polar no controlo de um sistema de visão activa. Esta geometria, ao privilegiar as zonas próximas do centro da imagem, é particularmente adequada a tarefas de seguimento de alvos móveis. Algoritmos usualmente aplicados em imagens cartesianas, como a correlação e o fluxo óptico, são aplicados com sucesso a imagens log-polar e com vantagens evidentes ao nível do tempo de processamento e dos desempenhos alcançados.

É feito um esforço no sentido de estudar e compreender a relação existente entre o processamento de imagens cartesianas e imagens com resolução não uniforme. Esta relação é estabelecida de um modo formal para o caso dos algoritmos de correlação. Análises semelhantes poderão ser feitas para outros algoritmos, e relacionadas directamente com a resolução das imagens. São também descritas formas sistemáticas de efectuar a discretização de imagens cartesianas de acordo com as transformações de coordenadas pretendidas.

Outro aspecto importante desta tese refere-se à inspiração nos sistemas de visão biológicos para motivar metodologias de projecto do sistema de controlo de seguimento. Em particular são estudados os tipos de movimentos oculares que cooperam no comportamento de seguimento e os estímulos que os despoletam no sistema visual humano. A partir deste estudo foram definidos movimentos e estímulos visuais básicos que se revelaram bastante úteis no sistema desenvolvido.

Em especial para a cabeça robótica Medusa, mas também para vários outros sistemas com estrutura semelhante, é estabelecida uma geometria de fixação própria que facilita a decomposição dos movimentos oculares em componentes separáveis e simplifica a análise da cinemática do sistema. São também apresentados métodos de projecto de controladores dinâmicos que têm em conta um modelo linear simplificado do sistema e os atrasos computacionais introduzidos pelas rotinas de processamento, e que são posteriormente integrados numa arquitectura conjunta.

1.5 Organização da tese

No Capítulo 2 apresenta-se o sistema robótico utilizado como base experimental nesta tese. O problema da fixação e seguimento de alvos móveis é relacionado com o controlo da posição do ponto de fixação no espaço. O estudo de alguns aspectos do sistema visual humano permite motivar uma decomposição para os movimentos oculares que simplifica o controlo do ponto de fixação. A cinemática do sistema é analisada de modo a avaliar como o ponto de fixação visual no espaço se desloca em função dos movimentos dos graus de liberdade, e como gerar estes movimentos de acordo com a decomposição encontrada. Ainda neste capítulo introduz-se o modelo de formação de imagem utilizado e aspectos relacionados com a binocularidade do sistema.

O Capítulo 3 é dedicado ao estudo de imagens com resolução não uniforme. Este estudo é motivado pela constituição não uniforme da retina humana e de outros seres vivos. São definidos alguns conceitos relacionados com a resolução de imagens contínuas e discretas, em que se assume a utilização de transformações de coordenadas para representar os processos que dão origem a essas imagens. A partir destes conceitos é possível estabelecer métodos de criação de imagens discretas relacionadas com uma determinada transformação de coordenadas. É dado especial ênfase à transformação log-polar, que será utilizada nas restantes fases do trabalho e, para esta transformação são definidos parâmetros úteis no dimensionamento das grelhas de discretização.

Com base na decomposição encontrada para os movimentos oculares e estímulos visuais relacionados, os Capítulos 4 e 5 descrevem os comportamentos de vergência e de seguimento desenvolvidos. Continua a ser dada especial atenção à utilização de imagens log-polar, em particular ao modo como, a partir destas imagens, obter estímulos apropriados para o controlo dos processos. Medidas de correlação, utilizadas em vergência, e medidas de posição e de velocidade do alvo, utilizadas no seguimento, beneficiam da geometria log-polar. São ainda ilustradas formas sistemáticas de adaptar algoritmos do domínio cartesiano para o log-polar e relacionar o desempenho nos dois casos.

A integração dos comportamentos de vergência e de seguimento é discutida no Capítulo 6. É visto como a partir da fusão binocular proporcionada pelo comportamento de vergência é possível discriminar as zonas da imagem que pertencem ao alvo de inte-

resse. Só nestas zonas da imagem é que se justifica obter os estímulos visuais necessários para o seguimento. Relativamente a este processo de segmentação é também avaliada a influência da utilização das imagens log-polar. É apresentada uma arquitectura para a coordenação dos dois comportamentos, em que o aspecto fundamental prende-se com a prioridade atribuída aos movimentos de vergência – só em situações correctas de fusão binocular se torna fiável obter os estímulos necessários para o seguimento.

Os resultados de diversas experiências efectuadas na cabeça robótica Medusa são apresentados no Capítulo 7. Evidencia-se o desempenho do sistema com objectos de várias formas e com diferentes tipos de movimento. A avaliação da precisão do seguimento é feita de uma forma qualitativa para objectos com movimento conhecido e para objectos que se movem livremente dentro do espaço visual.

Finalmente, no Capítulo 8 são apresentadas as conclusões do trabalho desenvolvido e estabelecem-se direcções de pesquisa em trabalho futuro.

2. Um observador artificial

Nesta tese é utilizado um sistema robótico de visão binocular. A suas capacidades motoras e perceptuais são analisadas neste capítulo. É sugerida uma decomposição para o movimento dos graus de liberdade do sistema robótico e estímulos visuais utilizados no seu controlo, motivada pelo estudo dos movimentos oculares de sistemas visuais biológicos, em especial aqueles que cooperam no processo de seguimento de alvos móveis.

2.1 A cabeça robótica Medusa

O recente interesse na construção de dispositivos mecânicos de estrutura antropomórfica (*cabeças robóticas*) capazes de simular os graus de liberdade principais do sistema visual do ser humano, constitui um passo importante no sentido da criação de um observador artificial [22]. A cabeça robótica **Medusa**, construída neste âmbito, constitui a base experimental para grande parte deste trabalho (ver Figura 2.1).



Fig. 2.1: A cabeça robótica Medusa.

É composta por um sistema *stereo*, com câmaras motorizadas independentemente, o que lhe confere a capacidade de alterar o ângulo formado pelos eixos ópticos e assim variar a profundidade do ponto de fixação no espaço. As câmaras são montadas sobre suportes comuns também motorizados que executam os movimentos de elevação (*tilt*) e rotação horizontal (*pan*), permitindo variar arbitrariamente a direcção de observação no campo visual.

Ao todo, a cabeça robótica Medusa contém 4 graus de liberdade motorizados que lhe possibilitam fixar qualquer ponto no espaço envolvente. Manualmente, as câmaras podem ainda ser ajustadas de modo a que os seus eixos de rotação e o eixo de *tilt* intersectem os centros ópticos, simplificando a cinemática do ponto de fixação. A distância entre as câmaras (distância interocular) também pode ser ajustada para alterar o espaço de trabalho, em termos da gama de profundidades utilizáveis.

A abordagem antropomórfica estende-se também às especificações dinâmicas da cabeça robótica. As velocidades máximas atingíveis são comparáveis com as do sistema visual humano (cerca de 180° s^{-1}) e as acelerações máximas excedem os $1000^\circ \text{ s}^{-2}$ [61].

O sistema de controlo oculomotor desenvolvido nesta tese é também baseado em algumas características do sistema visual humano, principalmente no que respeita aos movimentos oculares e estímulos visuais que são envolvidos no comportamento de fixação de atenção.

2.2 Movimentos oculares

Muitas tarefas visuais surgem da integração de diversos movimentos oculares simples. Na perspectiva do controlo oculomotor, interessa considerar os movimentos oculares básicos como os elementos atómicos na constituição do comportamento visual. Existem seis movimentos oculares básicos [53] no sistema visual humano:

- [*saccades*] – **movimento sacádico**: redireccionamento da direcção de observação de uma forma rápida e “cega” (sem recolha de informação visual enquanto decorre); são os únicos movimentos que podemos controlar voluntariamente;

- [*smooth pursuit*] – **movimento de seguimento lento**: movimento controlado por retroacção visual que só ocorre na presença de um estímulo de velocidade (que se revela através de um escorregamento do alvo na retina);
- [*vergence*] – **movimento de vergência**: alteração do ângulo formado pelos eixos ópticos, controlada visualmente, de modo a compensar os movimentos do alvo em profundidade;
- [*vestibular nystagmus*] – **nistagmo vestibular**: reflexo ocular que compensa o movimento próprio do observador por utilização de informação proveniente do sistema de equilíbrio (localizado no ouvido interno);
- [*optokinetic nystagmus*] – **nistagmo opto-cinético**: reflexo ocular que tenta estabilizar a imagem na retina quando grande parte do campo visual se move de forma coerente;
- [*physiological nystagmus*] – **nistagmo fisiológico**: pequenos movimentos de alta frequência, sempre presentes durante fixação, que têm como objectivo habilitar a percepção da informação luminosa.

Um aspecto importante em todos estes movimentos é o facto de serem neurologicamente separados. Para o projecto do observador artificial, isto motiva uma decomposição lógica dos movimentos oculares. Cada um pode ser desenvolvido individualmente, e posteriormente integrado no sistema completo, de forma adequada.

Porém, no nosso caso alguns destes movimentos não têm interesse prático. Por exemplo, o nistagmo fisiológico tem uma função muito específica relacionada com a forma como os nossos olhos percebem a luminosidade; utilizando câmaras comuns, este movimento deixa de ser necessário. O nistagmo vestibular, seria interessante considerar. No entanto, requer a utilização de dispositivos sensoriais complexos (por exemplo acelerómetros) e não será aplicado no presente trabalho.

2.2.1 Movimentos oculares e estímulos visuais na fixação

O comportamento de fixação é constituído principalmente por movimentos oculares involuntários. Movimentos de vergência, de seguimento lento e alguns movimentos sacádicos

cooperam no seu funcionamento [21].

Os movimentos de vergência são estimulados principalmente por indícios de disparidade retinal. Isto acontece quando um ponto do espaço têm projecções em localizações diferentes nas duas retinas, indicando que o alvo se encontra a uma profundidade diferente da que se encontra sob fixação. Os movimentos de vergência tendem a anular este erro através da rotação dos olhos em sentidos contrários.

Os movimentos de seguimento lento são responsáveis pelo seguimento do alvo em direcções laterais à direcção de observação, e utilizam estímulos de velocidade. Para isso tentam anular a velocidade do alvo na retina, designada por **escorregamento retinal**.

Também utilizados no seguimento lateral do alvo, mas desta vez com estímulos de posição retinal, são os movimentos sacádicos involuntários. Estes movimentos também são chamados de **micro-sacádicos**, pelas suas pequenas amplitudes quando comparadas com as dos movimentos sacádicos voluntários. O seu principal objectivo é o de corrigir pequenos erros que surgem durante os movimentos de seguimento lento e que se devem a movimentos mais bruscos do alvo ou a acumulação de erros ao longo do tempo. Têm um carácter discreto no tempo, sendo despoletados quando o erro de posição do alvo relativamente ao centro da retina excede certos valores. Quando isso acontece deslocam rapidamente a posição dos olhos e inibem temporariamente os movimentos de seguimento lento, processo designado por **supressão sacádica**.

Após esta descrição sucinta dos principais movimentos oculares constituintes da fixação visual, iremos em seguida analisar os movimentos que podemos efectuar com a cabeça robótica Medusa. Estamos principalmente interessados em proceder a uma decomposição dos movimentos em duas classes: compensação de deslocamentos do alvo em profundidade e compensação dos deslocamentos laterais. Esta decomposição é baseada no que ocorre no comportamento humano de fixação: os movimentos de vergência pertencem à primeira classe enquanto os outros executam o segundo tipo de movimentos.

2.3 Medusa: aspectos geométricos

A cabeça robótica Medusa tem 4 graus de liberdade motorizados implementados por juntas de revolução, duas juntas relativas às câmaras, uma junta de *pan* e uma junta

de *tilt*. A posição angular de cada junta é representada por intermédio de um ângulo, denominado por ângulo de junta, com a seguinte correspondência:

- θ_d — câmara direita;
- θ_e — câmara esquerda;
- θ_t — *tilt*;
- θ_p — *pan*.

A Figura 2.2 representa de uma forma esquemática os graus de liberdade da cabeça robótica Medusa, onde podemos avaliar qualitativamente a influência da alteração de cada um dos ângulos de junta.

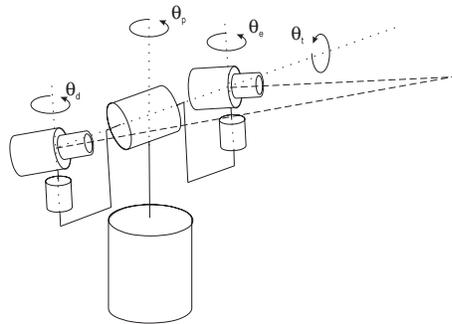


Fig. 2.2: Graus de liberdade da cabeça robótica Medusa.

As câmaras são ajustadas manualmente de tal forma que possam rodar em torno de eixos paralelos e que os seus eixos ópticos se intersectem algures no espaço. Para além disso, e para simplificar a cinemática do sistema, o eixo da junta de *tilt* é feito passar pelos centros ópticos das câmaras. O eixo da junta de *pan* é perpendicular ao eixo de *tilt* e é feito passar pelo ponto médio da distância entre as câmaras, por ajuste manual da posição destas.

O ponto de intersecção dos eixos ópticos de cada câmara é denominado por **ponto de fixação**. Os movimentos das câmaras deslocam o ponto de fixação ao longo do plano definido pelos eixos ópticos, a que chamaremos **plano de vergência**. Os movimentos da junta de *tilt* permitem alterar o ângulo entre este plano e o plano horizontal, provocando

o deslocamento do ponto de fixação segundo circunferências verticais, e os movimentos da junta de *pan* deslocam o ponto de fixação segundo circunferências horizontais.

Com uma geometria deste tipo é teoricamente possível fixar qualquer ponto do espaço envolvente, dentro dos limites permitidos para as juntas. Para além disso, existe redundância na atingibilidade de pontos no espaço, resultante da existência de quatro graus de liberdade para alcançar um ponto num espaço de dimensão 3: configurações diferentes podem dar origem a um mesmo ponto de fixação espaço. Como veremos em seguida, este facto pode ser simplificado restringindo os movimentos permitidos para as juntas das câmaras, resultando em correspondências unívocas entre pontos 3D e ângulos de junta.

2.3.1 Geometria binocular

Para ângulos de *pan* e *tilt* constantes, o ponto de fixação no espaço fica restrito ao plano de vergência. Pretende-se analisar como se move o ponto de fixação neste plano em função da rotação das câmaras esquerda e direita, de ângulos θ_e e θ_d respectivamente.

Considere-se o sistema de eixos $\{X, Z\}$ definido sobre o plano de vergência, representado na Figura 2.3.

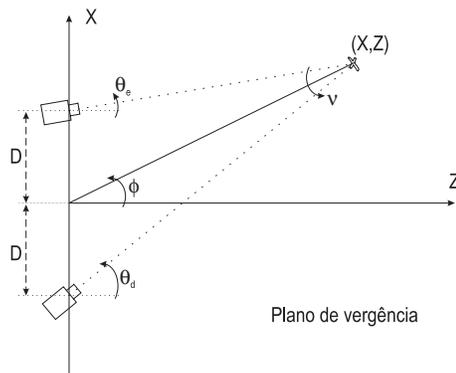


Fig. 2.3: Geometria binocular no plano de vergência.

O eixo X é definido pela recta que passa pelos centros ópticos das duas câmaras (**recta interocular**) e $2D$ é a **distância interocular**. O ponto de fixação encontra-se localizado num ponto genérico de coordenadas (X, Z) . Ao eixo que une a origem das coordenadas com o ponto de fixação chama-se **eixo ciclópico** e ϕ é o **ângulo ciclópico**. Ao ângulo formado pelos eixos ópticos (ν) chamaremos **ângulo de vergência**. Estes dois ângulos

são dados por:

$$\begin{aligned}\tan \phi &= \frac{\tan \theta_d + \tan \theta_e}{2} \\ \nu &= \theta_d - \theta_e\end{aligned}\tag{2.1}$$

Para a geometria dada podem se estabelecer as seguintes relações:

$$\begin{cases} \tan \theta_e = \frac{X-D}{Z} \\ \tan \theta_d = \frac{X+D}{Z} \end{cases}$$

Por simples manipulação algébrica obtém-se a posição do ponto de fixação em função dos ângulos das câmaras:

$$\begin{cases} X = D \frac{\tan \theta_d + \tan \theta_e}{\tan \theta_d - \tan \theta_e} \\ Z = \frac{2D}{\tan \theta_d - \tan \theta_e} \end{cases}$$

As expressões anteriores relacionam a posição do ponto de fixação com os ângulos das câmaras tomados como entidades separadas. No entanto, no sistema visual humano existe uma grande dependência interocular. Mesmo no caso simples considerado, em que se limitam os deslocamentos do ponto de fixação ao plano de vergência, não podemos considerar um movimento ocular entre dois pontos do espaço como uma rotação única e independente dos olhos que os levam desde a posição inicial até à posição desejada.

Como já vimos, duas classes de movimentos cooperam neste sentido. Uma, relacionada com os movimentos de vergência, provoca rotações dos olhos em sentidos opostos de modo a deslocar o ponto de fixação em profundidade. A outra, provoca rotações oculares com o mesmo sentido para transladar lateralmente o ponto de fixação. Embora as rotações oculares atrás referidas não possam ser tomadas como perfeitamente simétricas no primeiro caso, ou perfeitamente idênticas no segundo caso, uma decomposição deste tipo permite ilustrar a geometria subjacente a estes mecanismos. Por outro lado seria desejável encontrar condições que permitissem a decomposição do movimento ocular em duas componentes:

- deslocações puras em profundidade, ao longo do eixo ciclópico;
- deslocações laterais puras, ortogonais ao eixo ciclópico.

Consideremos então:

$$\begin{cases} \theta^+ = \frac{\theta_d + \theta_e}{2} \\ \theta^- = \frac{\theta_d - \theta_e}{2} \end{cases} \Leftrightarrow \begin{cases} \theta_e = \theta^+ - \theta^- \\ \theta_d = \theta^+ + \theta^- \end{cases}$$

em que θ_+ representa a rotação comum das duas câmaras e θ_- representa a rotação diferencial. Em termos destes ângulos o ponto de fixação vem dado por:

$$\begin{cases} X = D \frac{\sin(\theta^+) \cos(\theta^+)}{\sin(\theta^-) \cos(\theta^-)} \\ Z = D \frac{\cos(\theta^+ + \theta^-) \cos(\theta^+ - \theta^-)}{\sin(\theta^-) \cos(\theta^-)} \end{cases} \quad (2.2)$$

Analisando a influência dos ângulos θ^+ e θ^- independentemente, obtemos trajectórias para o ponto de fixação da forma apresentada na Figura 2.4:

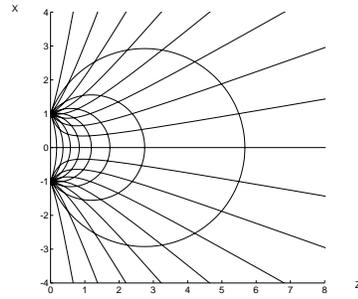


Fig. 2.4: Trajectórias do ponto de fixação devidas a rotações idênticas ou simétricas das câmaras, em diversas configurações.

- Aplicando variações de ângulo idênticas às duas câmaras, o ponto de fixação descreve trajectórias circulares que passam pelos centros ópticos das câmaras e pela intersecção dos eixos ópticos. Estas linhas são habitualmente designadas por círculos de Vieth–Muller [47].
- A aplicação de variações de ângulo simétricas, provoca trajectórias que são hipérbolas rectangulares centradas no ponto médio da distância interocular. Para $\theta^+ = 0$, então a trajectória degenera numa recta e coincide com o eixo ciclópico.

Na geometria considerada, pode-se estabelecer uma relação linear directa entre o ângulo de vergência e o ângulo de rotação diferencial ($\nu = 2\theta^-$). O mesmo não acontece entre o ângulo ciclópico e o ângulo de rotação comum. Expressando (2.1) como função de θ_+ e θ_- , temos:

$$\tan \phi = \frac{\sin(\theta^+) \cos(\theta^+)}{\cos(\theta^+ + \theta^-) \cos(\theta^+ - \theta^-)}$$

pelo que se conclui que $\phi = \theta^+$ só para os casos limites de $\theta^- = 0$ (fixação no infinito) ou $\theta^+ = 0$. Esta última condição, correspondente à situação $\theta_e = -\theta_d$, é denominada por **fixação frontal**, e simplifica a análise do movimento nas duas componentes separáveis pretendidas. Em rigor, só nesta situação é que rotações simétricas das câmaras correspondem a deslocações puras em profundidade e rotações idênticas provocam translações localmente ortogonais ao eixo ciclópico.

2.3.2 Utilização de coordenadas esféricas

Consideremos agora também rotações das juntas de *pan* e de *tilt*, de ângulos θ_p e θ_t , respectivamente. Como vimos atrás, a situação de fixação frontal corresponde ao caso em que podemos proceder à separação dos movimentos oculares em componentes ortogonais. Para além disso, esta situação permite relacionar com simplicidade a geometria completa da cabeça robótica com as usuais coordenadas esféricas (ρ, ϕ, γ):

$$\begin{cases} X = \rho \cos \gamma \sin \phi \\ Y = \rho \sin \gamma \\ Z = \rho \cos \gamma \cos \phi \end{cases}$$

que se encontram representadas na Figura 2.5.

Consideremos uma situação de fixação frontal e um referencial com origem no ponto médio da linha interocular, como no diagrama da Figura 2.6.

É fácil reconhecer que a distância entre a origem do referencial e o ponto de fixação é medida ao longo do eixo ciclópico e corresponde ao raio (coordenada ρ) do sistema de coordenadas esférico. Esta distância apenas depende do ângulo de vergência das câmaras e é calculada particularizando a Equação (2.2):

$$\rho = D \cotg\left(\frac{\nu}{2}\right)$$

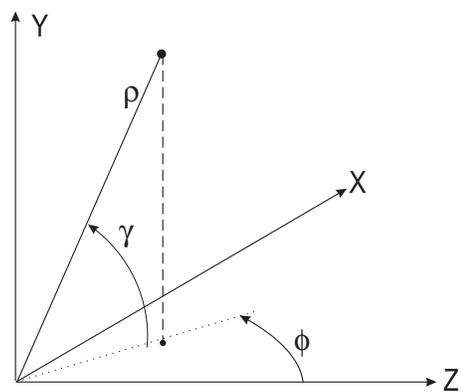


Fig. 2.5: Coordenadas esféricas.

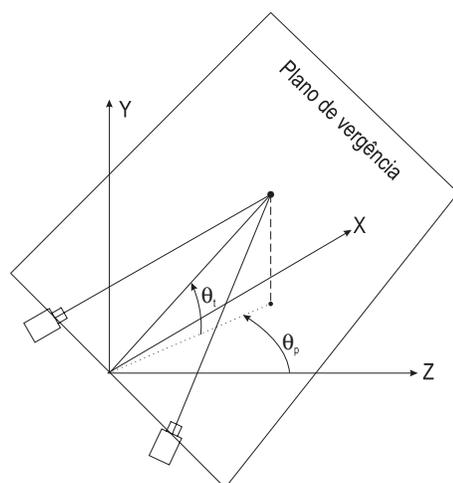


Fig. 2.6: O ponto de fixação sob a hipótese de fixação frontal.

Como na fixação frontal o eixo ciclópico é perpendicular à linha interocular, então os ângulos de *pan* e de *tilt* correspondem às coordenadas ϕ e γ .

Assim, admitindo fixação frontal, movimentos de *pan* e de *tilt* deslocam o ponto de fixação numa superfície esférica de raio ρ , em deslocamentos localmente ortogonais ao eixo ciclópico, excepto em casos singulares¹. As estes movimentos, por apenas alterarem a direcção de observação e não a profundidade do ponto de fixação, designaremos por **movimentos direccionais**.

Resumindo, a utilização de fixação frontal permite abordar a cinemática da cabeça robótica por utilização das coordenadas esféricas. Consequentemente, torna-se possível decompor os movimentos do ponto de fixação em duas componentes ortogonais:

- movimentos puros de profundidade ao longo do eixo ciclópico, gerados por rotação simétrica das câmaras e relacionados com a coordenada ρ ;
- movimentos direccionais puros; horizontais quando gerados pela junta de *pan* e verticais quando gerados pela junta de *tilt*, como se pode ver na Figura 2.7. Os ângulos θ_p e θ_t são iguais aos ângulos do sistema de coordenadas esféricas, ϕ e γ respectivamente.

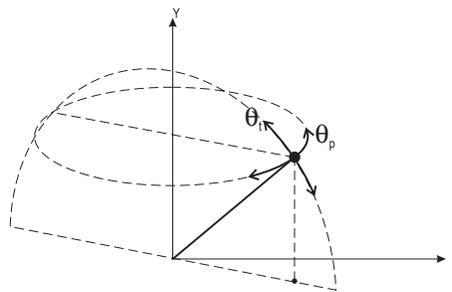


Fig. 2.7: Movimentos fronto-paralelos efectuados por rotações de *pan* e de *tilt*.

2.4 Formação de imagem e disparidade binocular

Até aqui analisámos a forma como o observador artificial pode controlar o ponto de fixação visual por alteração da sua configuração geométrica.

¹por exemplo, se $\theta_t = 90^\circ$ então nenhum movimento é gerado por rotação da junta de *pan*.

Nesta secção vamos ver como a percepção do mundo é influenciada pela posição dos objectos no espaço visual. Em particular, analisa-se como os pontos se projectam no plano de imagem, gerando estímulos de disparidade binocular e introduz-se a noção de horóptero binocular.

2.4.1 Modelo de formação de imagens

O modelo de formação de imagens por projecção perspectiva [33] é o modelo aplicado à maior parte dos dispositivos ópticos existentes devido à sua simplicidade e boa aproximação à realidade. O modelo de projecção perspectiva encontra-se representado na Figura 2.8.

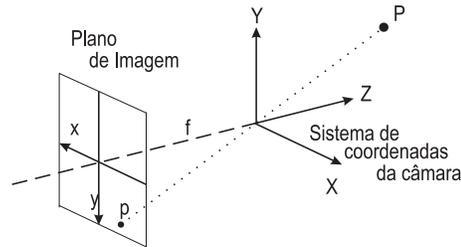


Fig. 2.8: Modelo de formação de imagem por projecção perspectiva.

Este modelo relaciona as coordenadas de um ponto do mundo exterior (X, Y, Z) com as coordenadas da sua projecção no plano de imagem (x, y) , da seguinte forma:

$$x = f \frac{X}{Z}; y = f \frac{Y}{Z}$$

em que f é a distância focal da lente. Sem perda de generalidade, daqui em diante este parâmetro será considerado unitário.

A projecção de um ponto do espaço nos planos de imagem de duas câmaras fica assim completamente determinada, desde que esse ponto esteja referido a sistemas de coordenadas solidários com as câmaras. Na próxima secção determinam-se as projecções nos planos de imagem de um ponto representado num referencial comum às duas câmaras. A diferença entre essas projecções, designada por disparidade, constitui uma pista visual importante no controlo dos movimentos de vergência.

2.4.2 Disparidade

A disparidade é definida como a distância entre as projecções de um ponto do espaço nos planos de imagem de duas câmaras (ver Figura 2.9).

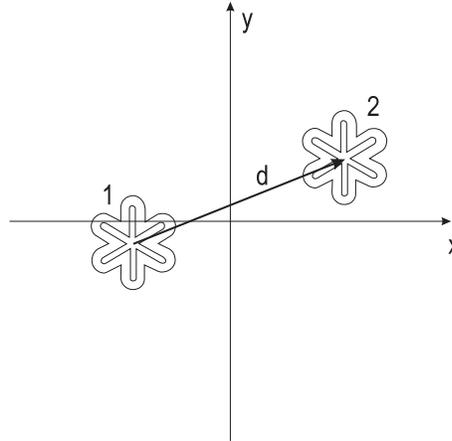


Fig. 2.9: Se imaginarmos as retinas sobrepostas, a distância entre a posição do objecto na retina 1 e na retina 2 é denominada por disparidade retinal (d).

Se as coordenadas de dois pontos correspondentes nos planos de imagem esquerdo e direito forem (x_e, y_e) e (x_d, y_d) respectivamente, então o valor da disparidade vem dado por:

$$d = \sqrt{(x_e - x_d)^2 + (y_e - y_d)^2}$$

onde os termos $(x_e - x_d)$ e $(y_e - y_d)$ são as disparidades horizontal e vertical, respectivamente.

No caso particular da cabeça robótica Medusa, vamos adoptar a hipótese de fixação frontal que introduz vantagens ao nível da simplificação geométrica e da decomposição dos movimentos oculares. Considere-se a configuração geométrica apresentada na Figura 2.10, em que o referencial com origem no ponto médio da distância entre as câmaras é designado por **referencial ciclópico** $\{C\}$, uma vez que o seu eixo coordenado Z_c coincide com o eixo ciclópico do sistema binocular. Introduzem-se também os referenciais $\{E\}$ e $\{D\}$ que se movem solidariamente com as câmaras esquerda e direita respectivamente.

A partir da análise geométrica do problema e admitindo a projecção perspectiva no processo de formação das imagens, podemos concluir que um ponto (X, Y, Z) , descrito no

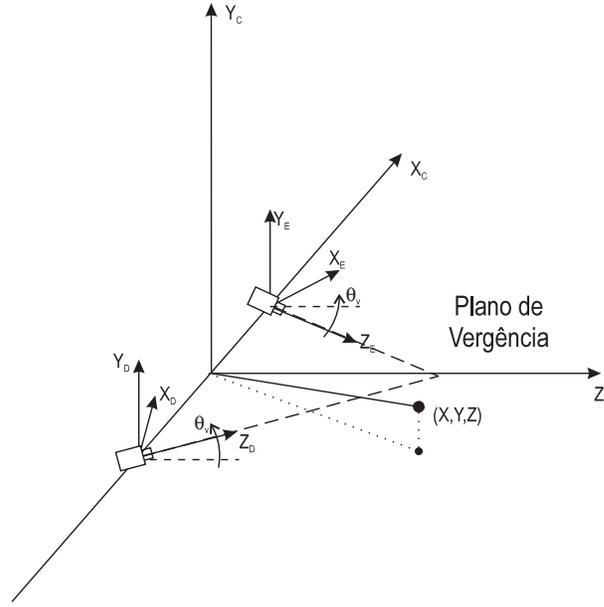


Fig. 2.10: Sistemas de coordenadas utilizados para o estudo da disparidade binocular.

referencial ciclópico, projecta-se nos planos de imagem nas seguintes coordenadas:

$$\left\{ \begin{array}{l} x_e = \frac{(X+D) \cos \theta_v - Z \sin \theta_v}{(X+D) \sin \theta_v + Z \cos \theta_v} \\ y_e = \frac{Y}{(X+D) \sin \theta_v + Z \sin \theta_v} \end{array} \right. \quad \text{e} \quad \left\{ \begin{array}{l} x_d = \frac{(X-D) \cos \theta_v + Z \sin \theta_v}{-(X-D) \sin \theta_v + Z \cos \theta_v} \\ y_d = \frac{Y}{-(X-D) \sin \theta_v + Z \sin \theta_v} \end{array} \right.$$

Assim, as disparidades horizontais e verticais vêm dadas por:

$$\left\{ \begin{array}{l} x_e - x_d = -2 \frac{(X^2 + Z^2 - D^2) \sin \theta_v \cos \theta_v + ZD(\sin^2 \theta_v - \cos^2 \theta_v)}{(Z \cos \theta_v + D \sin \theta_v)^2 - X^2 \sin^2 \theta_v} \\ y_e - y_d = 2 \frac{XY \sin \theta_v}{(Z \cos \theta_v + D \sin \theta_v)^2 - X^2 \sin^2 \theta_v} \end{array} \right. \quad (2.3)$$

De um modo geral estas expressões são de difícil interpretação. No entanto, os estímulos de disparidade são especialmente utilizados no controlo dos movimentos de vergência, que deslocam o ponto de fixação ao longo do eixo ciclópico. Restringindo a análise a esta direcção ($X = 0, Y = 0$), ficamos com:

$$\left\{ \begin{array}{l} x_e = \frac{D \cos \theta_v - Z \sin \theta_v}{D \sin \theta_v + Z \cos \theta_v} \\ y_e = 0 \end{array} \right. \quad \text{e} \quad \left\{ \begin{array}{l} x_d = -\frac{D \cos \theta_v - Z \sin \theta_v}{D \sin \theta_v + Z \cos \theta_v} \\ y_d = 0 \end{array} \right.$$

ou seja, só existe disparidade horizontal e as projecções do ponto nos planos de imagem

surtem em localizações simétricas. A amplitude da disparidade vem dada por:

$$d = 2 \frac{D \cos \theta_v - Z \sin \theta_v}{D \sin \theta_v + Z \cos \theta_v}$$

A Figura 2.11 apresenta o valor da disparidade como função do ângulo de vergência, para diversos pontos ao longo do eixo ciclópico. Apesar da disparidade ser uma função não linear da distância do objecto e do ângulo de vergência, é de destacar o facto de esta evoluir de uma forma quase linear ao longo de um largo intervalo de ângulos de vergência, e com um declive muito semelhante para as diversas posições do alvo. Este facto permitirá a aplicação de um sistema de controlo de posição linear ao movimento de vergência para uma larga zona de funcionamento.

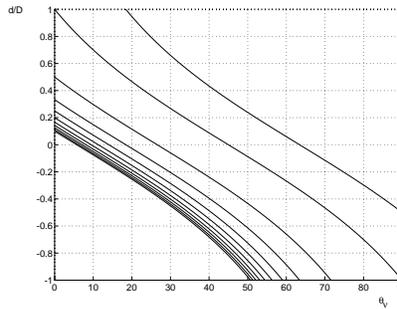


Fig. 2.11: Relação entre disparidade e ângulo de vergência para diversos ponto de fixação ao longo do eixo ciclópico (entre 0.25 e 5 vezes a distância interocular).

2.4.3 Horóptero geométrico

Outra questão interessante no estudo da disparidade binocular consiste em procurar os pontos do espaço que produzem disparidade nula para uma determinada configuração binocular. A este conjunto de pontos chama-se **horóptero² geométrico** [23]. Se o objecto se mover ao longo desses pontos então não estimula quaisquer movimentos de vergência.

Igualando a zero a Equação (2.3), que exprime o valor da disparidade como função da

²do inglês, *horopter*.

posição do ponto e do ângulo de vergência em fixação frontal, obtemos:

$$\begin{cases} Y = 0 \\ X^2 + Z^2 - D^2 - DZ \frac{\cos^2 \theta_v - \sin^2 \theta_v}{\cos \theta_v \sin \theta_v} = 0 \end{cases}$$

o que corresponde a uma circunferência no plano de vergência, de raio R e centro em $(X = 0, Z = Z_0)$, com:

$$\begin{cases} R = \left| \frac{D}{2 \cos \theta_v \sin \theta_v} \right| \\ Z_0 = \frac{D}{2} \frac{\cos^2 \theta_v - \sin^2 \theta_v}{\cos \theta_v \sin \theta_v} \end{cases}$$

Na Figura 2.12 pode-se observar que estas circunferências passam pelo ponto de fixação original no eixo ciclópico e pelos centros ópticos das câmaras.

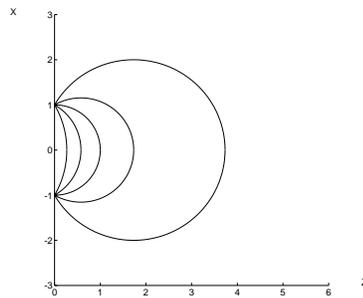


Fig. 2.12: O horóptero geométrico é formado por circunferências que passam pelo ponto de fixação e pelos centros ópticos das câmaras.

Todos os pontos que se encontram nestas localizações do espaço ocupam posições idênticas nas duas retinas, mas distintas de ponto para ponto. Pontos sob fixação projectam-se no centro das retinas mas, no caso geral, não é forçoso que isso aconteça.

Que movimentos oculares são necessários efectuar para compensar estes erros de posição? Relembremos os círculos de Vieth-Muller (ver Secção 2.3.1), que descrevem a trajectória do ponto de fixação sob rotações idênticas das câmaras. Comparando as Figuras 2.12 e 2.4 constatamos que a forma do horóptero geométrico é idêntica à dos círculos de Vieth-Muller. Assim, movimentos comuns nas duas câmaras permitirão, por si só, compensar os erros de posição dos alvos no horóptero. No nosso caso, em que movimentos das câmaras apenas se destinam a efectuar movimentos de vergência, essa

compensação é efectuada por uma combinação de movimentos de *pan* e de vergência. Localmente, apenas movimentos de *pan* serão gerados. No entanto, como este grau de liberdade não desloca o ponto de fixação no horóptero, dará origem a uma disparidade não nula e a uma reacção por parte do movimento de vergência.

Veremos num dos próximos capítulos que os pontos pertencentes ao horóptero constituem pistas importantes no controlo dos movimentos de seguimento.

3. Retina não uniforme.

Grande parte deste trabalho baseia-se na utilização de imagens de resolução não uniforme num sistema de visão por computador. Neste capítulo focam-se aspectos de representação e de construção de imagens deste tipo. É dado especial destaque para a geometria log-polar, que se inspira na distribuição dos fotoreceptores na retina humana.

3.1 Imagens com resolução não uniforme.

3.1.1 Resolução

As questões de como representar e como obter imagens de resolução não uniforme com as propriedades desejadas são abordadas nesta secção. Um primeiro passo neste sentido consiste na definição do conceito de **resolução de uma imagem**. Intuitivamente este conceito está relacionado com a quantidade de informação pictórica utilizada para representar uma determinada zona do espaço visual: uma grande quantidade de informação significa uma elevada resolução na representação dessa zona. A próxima definição estabelece esta relação para cada zona de uma imagem.

Definição 3.1: Resolução de uma imagem. Considere-se uma imagem I definida num conjunto A . Qualquer região da imagem ($B \subset A$), representa uma determinada zona do espaço visual (E). Define-se resolução da imagem I na região B como a relação entre a quantidade de informação existente na região B e a quantidade de informação contida na zona do espaço visual representado.

$$R(B) = \frac{Q(B)}{Q(E)}$$

Para que a definição anterior adquira significado é necessário estabelecer formas de medir a quantidade de informação contida numa região de uma imagem e a quantidade de informação contida numa zona do espaço visual representado. Para responder à segunda questão iremos definir o conceito de representação padrão do espaço visual, recorrendo à utilização de *imagens contínuas de resolução uniforme*. No entanto, há que ter em conta que uma imagem contínua obtida por um dispositivo de aquisição tenta representar, através de uma projecção num plano bidimensional, o ambiente envolvente de dimensão 3, havendo necessariamente perda de informação. Portanto, a preservação da topologia do ambiente (dimensão de objectos e distâncias entre objectos), não é possível no caso geral. A definição terá que cingir-se a uma aproximação desta realidade através de uma transformação projectiva.

Definição 3.2: Imagem contínua de resolução uniforme como padrão de representação do espaço visual. Se o modelo de projecção de perspectiva for válido para a aquisição de uma imagem contínua I definida no conjunto $A \subset R^2$, então diz-se que:

- I é uma **imagem contínua de resolução uniforme** (ou **imagem cartesiana**);
- I é um **padrão de representação do espaço visual**, ou seja, a quantidade de informação contida numa região da imagem ($B \subset A$) e a quantidade de informação existente no espaço visual representado (E) são idênticas.

Portanto, para determinar a resolução de uma imagem numa dada região teremos que calcular a quantidade de informação existente nessa região e relacioná-la com a quantidade de informação contida no padrão de representação correspondente. O problema resume-se a calcular a quantidade de informação numa região de uma imagem contínua ou discreta.

3.1.2 Resolução de imagens contínuas.

A questão que resta esclarecer para poder determinar a resolução de uma imagem contínua é a de como calcular a quantidade de informação de uma região da imagem. Embora para as imagens discretas este conceito seja facilmente interpretado, uma vez que podemos relacionar a quantidade de informação com o número de amostras discretas existentes,

para as imagens contínuas a extensão do conceito não é tão simples. Não existindo elementos discretos como unidades de informação, há que convencionar uma medida da quantidade de informação contida numa zona de uma imagem contínua.

Definição 3.3: Área como medida da quantidade de informação. Seja uma imagem contínua $I(x, y)$, definida no conjunto $A \subset \mathbb{R}^2$. A quantidade de informação contida numa região $B \subset A$, ($Q(B)$) é dada pela área da região B :

$$Q(B) = \text{Área}(B) = \iint_B dx dy$$

De acordo com esta medida, e sendo I uma imagem contínua, a resolução de uma região B dessa imagem será dada simplesmente pela relação entre a sua área e a área da região E correspondente na imagem cartesiana, ou seja:

$$R(B) = \frac{\text{Área}(B)}{\text{Área}(E)}$$

A questão que agora se coloca é a de como determinar a região da representação padrão do espaço visual correspondente a uma região da imagem, ou seja, que tipo de mapeamento as relaciona. Relembremos que a representação padrão do espaço visual é obtida através do modelo de projecção de perspectiva para o sistema de aquisição de imagem, modelo que relaciona as coordenadas de um ponto do mundo exterior (X, Y, Z) com as coordenadas da sua projecção no plano de imagem (x, y) , através da Equação (2.4.1). Portanto, podemos considerar duas formas de obter imagens com resolução não uniforme: alteração da óptica do dispositivo de projecção de modo que o modelo de projecção de perspectiva já não seja válido (por exemplo, com lentes que introduzam distorção radial) ou transformação não linear do sistema de eixos do plano de imagem. Na maior parte dos casos de interesse prático, estas formas podem ser interpretadas como a aplicação de uma transformação de semelhança \mathbf{T} às coordenadas (x, y) da imagem. A definição seguinte estabelecerá como a utilização de mudanças de coordenadas permite representar imagens com resolução não uniforme e relacionar regiões dessas imagens com as suas correspondentes cartesianas.

Definição 3.4: Imagens e regiões isomorfas. As imagens $I(x, y)$ e $I'(w, z)$ dizem-se **isomorfas** se existir uma transformação de coordenadas $\mathbf{T} : A \rightarrow B$, bijetiva de $A \in \mathbb{R}^2$ em $B \in \mathbb{R}^2$ e diferenciável em A , tal que:

$$I'(w, z) = I[\mathbf{T}^{-1}(w, z)] \quad \text{e} \quad I(x, y) = I'[\mathbf{T}(x, y)]$$

Duas **regiões** $A' \subset A$ e $B' \subset B$ dizem-se **isomorfas** se:

$$B' = \mathbf{T}(A') \quad \Leftrightarrow \quad A' = \mathbf{T}^{-1}(B')$$

Segundo este raciocínio, se $I(x, y)$ representar uma imagem contínua de resolução uniforme (cartesiana) e $I'(w, z)$ uma imagem isomorfa obtida pela transformação de coordenadas $(w, z) = \mathbf{T}(x, y)$ e esta transformação for não linear, então $I'(w, z)$ será uma imagem com resolução não uniforme. O próximo teorema estabelece a relação entre a resolução de uma imagem e uma grandeza de forte expressão matemática – o *Jacobiano* da transformação de coordenadas.

Teorema 3.1: Resolução de uma imagem contínua. A resolução de uma imagem $I'(w, z)$ isomorfa a uma imagem de resolução uniforme $I(x, y)$ pela transformação de coordenadas $(w, z) = \mathbf{T}(x, y)$, é dada localmente pelo *Jacobiano* da transformação \mathbf{T} .

Demonstração:

Consideremos C_{xy} uma célula elementar contida no domínio de I , centrada no ponto (x, y) . Esta célula é mapeada, pela transformação \mathbf{T} , numa célula isomorfa no domínio de I' , C_{wz} , centrada no ponto (w, z) . Pelas Definições 3.1, 3.2 e 3.3, a resolução da imagem I' na célula C_{wz} é dada por:

$$R(C_{wz}) = \frac{\text{Área}(C_{wz})}{\text{Área}(C_{xy})} = \frac{\iint_{C_{wz}} dw dz}{\iint_{C_{xy}} dx dy}$$

Aplicando a mudança de coordenadas $(w, z) = \mathbf{T}(x, y)$ à expressão anterior, e tendo em conta que $\mathbf{T}^{-1}(C_{wz}) = C_{xy}$, ficamos com:

$$R[\mathbf{T}(C_{xy})] = \frac{\iint_{C_{xy}} |\mathbf{J}(x, y)| dx dy}{\iint_{C_{xy}} dx dy} = \frac{\iint_{C_{xy}} |\mathbf{J}(x, y)| dx dy}{\iint_{C_{xy}} dx dy}$$

em que \mathbf{J} é a matriz Jacobiana da transformação de coordenadas \mathbf{T} :

$$\mathbf{J}(x, y) = \begin{bmatrix} \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \\ \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \end{bmatrix}$$

Se fizermos as áreas das células elementares tender para valores infinitesimais ($C_{xy} \rightarrow 0$), teremos:

$$\lim_{C_{xy} \rightarrow 0} R[\mathbf{T}(C_{xy})] = \lim_{C_{xy} \rightarrow 0} \frac{\iint_{C_{xy}} |\mathbf{J}(x, y)| \, dx dy}{\iint_{C_{xy}} dx dy}$$

Quando a dimensão das células tende para 0, a região ocupada torna-se pontual, e o Jacobiano no integral pode ser substituído pelo seu valor na localização da célula. Assim sendo:

$$R[\mathbf{T}(x, y)] = |\mathbf{J}(x, y)|$$

q.e.d.

Como primeiro exemplo, vamos ilustrar a obtenção imagens de resolução não uniforme através da transformação polar.

Exemplo 3.1: Transformação polar. A transformação para coordenadas polares é dada por:

$$(\rho, \theta) = \mathbf{T}(x, y) = \left(\sqrt{x^2 + y^2}, \arctan\left(\frac{y}{x}\right) \right)$$

definida para $(x, y) \in R \setminus (0, 0)$.

Representando no plano (x, y) células elementares rectangulares, as células correspondentes no plano (ρ, θ) têm a forma representada na Figura 3.1. A área das células no plano polar diminui para a coordenada radial crescente.

O Jacobiano da transformação polar é dado por $|\mathbf{J}| = 1/\rho$, pelo que a resolução da imagem decresce pela proporção inversa de ρ . A Figura 3.2 apresenta curvas de nível do Jacobiano da transformação polar, representadas no plano cartesiano. Estas curvas identificam pontos de resolução constante e, quanto maior for a proximidade entre as curvas, maior é o gradiente de resolução. Na proximidade da origem a resolução tende para um valor infinito (note-se que a transformação não está definida na origem).

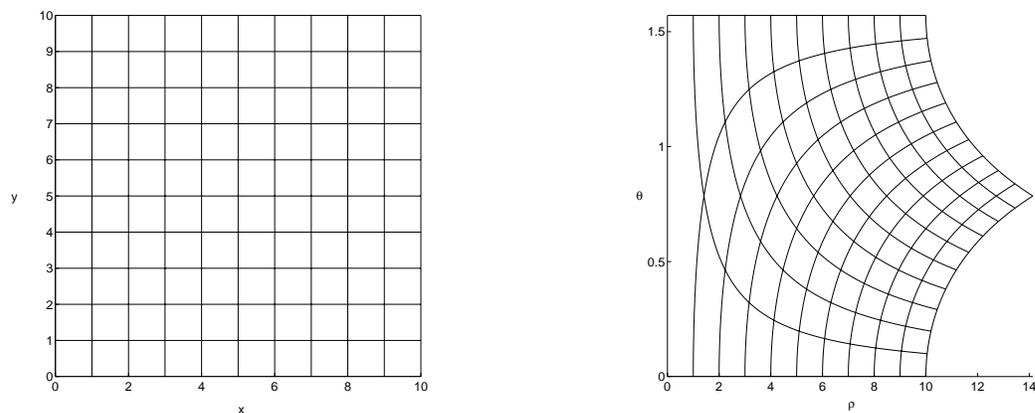


Fig. 3.1: Transformação de células elementares cartesianas (esquerda) para coordenadas polares (direita).

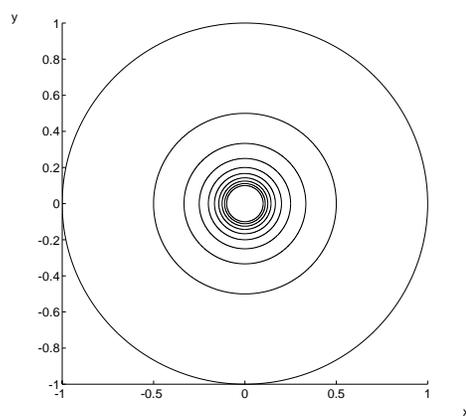


Fig. 3.2: Curvas de resolução constante para a transformação polar, representadas no plano cartesiano.

3.1.3 Resolução de imagens discretas.

Até aqui considerou-se a imagem definida num domínio contínuo. No entanto, os sistemas de visão artificial, devido à utilização generalizada dos computadores digitais, representam as imagens como sinais discretos bidimensionais. A maioria dos dispositivos de aquisição de imagem (por exemplo, tipo CCD - *charge coupled devices*) integram os valores da irradiância ao longo de células rectangulares de igual dimensão e o valor obtido é considerado como o nível de brilho no ponto central da célula. Com base neste motivo tecnológico vamos considerar a discretização de uma imagem contínua como o processo de subdividir a área da imagem em células arbitrárias (indexadas convenientemente), seguido do cálculo do valor médio da imagem em cada célula e da atribuição desse valor ao brilho da imagem discreta no ponto correspondente ao índice da célula.

Definição 3.5: Discretização de uma imagem contínua. Seja $I(x, y)$ a imagem contínua a discretizar, definida no conjunto $A \in R^2$. Considere-se um conjunto de células indexadas $G = \{C_{ij}\}$, denominado por **grelha de discretização**, sendo cada célula C_{ij} um sub-conjunto de A .

Define-se **discretização de $I(x, y)$ pela grelha de discretização G** como a imagem discreta $I^d(i, j)$, tal que:

$$I^d(i, j) = \frac{\int_{C_{ij}} I(x, y) \, dx dy}{\int_{C_{ij}} dx dy}$$

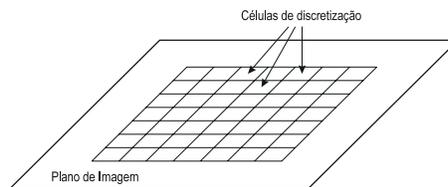


Fig. 3.3: Grelha de discretização habitualmente utilizada em CCD's.

Como já foi referido, os dispositivos de aquisição de imagem do tipo CCD, procedem à discretização de imagens contínuas através de grelhas compostas por células rectangulares dispostas uniformemente sobre a área de aquisição, como pode ser observado na Figura 3.3. A todas as grelhas constituídas por células de igual área e dispostas com igual

espaçamento, denominaremos por **grelhas uniformes**. Às discretizações de imagens efectuadas por este tipo de grelhas chamaremos **discretizações uniformes**.

O conceito de resolução para uma imagem discreta será definido, à semelhança do caso contínuo, como a relação entre a quantidade de informação utilizada para representar uma determinada realidade visual e a quantidade de informação contida no padrão de representação correspondente. Para isso há que convencionar a quantidade de informação contida numa amostra de uma imagem discreta, para a poder relacionar com a quantidade de informação da região da imagem cartesiana representada por essa amostra.

Definição 3.6: Elemento pictórico de uma imagem discreta como quantidade de informação. Cada amostra uma imagem discreta $I^d(i, j)$ é denominada por Elemento Pictórico¹ e representa uma quantidade de informação unitária, ou seja, pode ser considerado como uma região D_{ij} tal que:

$$\text{Área}(D_{ij}) = 1$$

Baseado nesta definição podemos estabelecer a forma de determinar a resolução de uma imagem discreta em cada ponto.

Resolução de uma imagem discreta. Seja $I^d(i, j)$ uma imagem obtida pela discretização de uma imagem contínua de resolução uniforme, $I(x, y)$, por uma grelha de discretização $G = \{C_{ij}\}$. De acordo com a Definição 3.6, cada ponto (i, j) , pode ser considerado como uma região D_{ij} , com área unitária. Pela Definição 3.5, cada ponto (i, j) representa uma região na imagem a discretizar subjacente à célula de discretização C_{ij} . Como, por hipótese, a imagem a discretizar é uma imagem contínua de resolução uniforme (padrão adoptado para a representação do espaço visual), então pelas Definições 3.1, 3.2 e 3.3, a resolução da imagem I^d é dada por:

$$R^d(D_{ij}) = \frac{\text{Área}(D_{ij})}{\text{Área}(C_{ij})}$$

Dado que a célula D_{ij} tem área unitária e representa unívocamente o ponto (i, j) ,

¹do inglês, *Picture Element*, ou *pixel*.

então:

$$R^d(i, j) = \frac{1}{\text{Área}(C_{ij})}$$

Exemplo 3.2: Grelha de discretização polar. A grelha de discretização polar obtém-se a partir da transformação de coordenadas polares para cartesianas:

$$(x, y) = \mathbf{S}(\rho, \theta) = (\rho \cos \theta, \rho \sin \theta)$$

definida para $(\rho, \theta) \in]0, +\infty[\times [0, 2\pi[$.

Cada célula unitária no plano (ρ, θ) é mapeada, através de \mathbf{S} , numa secção de anel circular no plano (x, y) . O mapeamento completo da grelha padrão resulta na grelha polar, representada na Figura 3.4, em que se admitem 10 passos de discretização em cada coordenada.

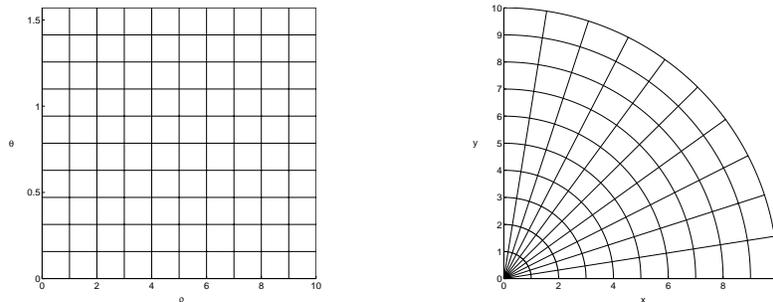


Fig. 3.4: Grelha de discretização uniforme no plano (ρ, θ) (esquerda) corresponde a uma grelha polar no plano cartesiano (direita).

Cada célula da grelha polar, C_{ij}^S , tem uma área dada por:

$$A_{ij} = \frac{\pi((i+1)^2 - i^2)}{N} = \frac{\pi(2i+1)}{N}$$

em que N representa o número de células na coordenada θ . A resolução de uma imagem discretizada através da grelha polar vem como:

$$R^d(i, j) = \frac{N}{\pi(2i+1)}$$

ou seja, é inversamente proporcional à coordenada ρ (definida aqui pelo índice i), tal como acontecia para as imagens contínuas obtidas pela transformação para coordenadas polares (ver Exemplo 3.1). Relembremos que a transformação para coordenadas polares foi definida como:

$$\mathbf{T}(x, y) = \left(\sqrt{x^2 + y^2}, \arctan\left(\frac{y}{x}\right) \right), \quad (x, y) \in R \setminus (0, 0)$$

Facilmente se verifica que $\mathbf{S} = \mathbf{T}^{-1}$.

O Teoremas 3.1 e o Exemplo 3.2 estabelecem uma importante analogia entre o Jacobiano de uma transformação de coordenadas e a área de células de uma grelha de discretização. Quando a transformação de coordenadas ou o processo de discretização são aplicados a imagens de resolução uniforme, a resolução da imagem obtida é determinada através do Jacobiano da transformação, no caso contínuo, ou da área das células de discretização, no caso discreto. Em seguida veremos como aprofundar esta analogia e através dela obter imagens de resolução não uniforme com as propriedades desejadas.

3.1.4 Aquisição de imagens discretas com resolução não uniforme

De um modo geral é possível obter imagens discretas com resolução não uniforme quer através de uma discretização uniforme de imagens contínuas de resolução não uniforme (*mapeamento contínuo*), quer através de uma discretização não uniforme de imagens contínuas de resolução uniforme (*mapeamento discreto*). Vimos atrás como é possível analisar a resolução de tais imagens. Pretendemos agora, sintetizar quando possível, imagens com uma resolução pretendida e relacionar os dois métodos de aquisição referidos.

Interessa-nos particularmente encontrar uma forma sistemática de construir grelhas de discretização que proporcionem através do *mapeamento discreto* resoluções semelhantes àsquelas obtidas pelo *mapeamento contínuo*. Para isso começamos por definir dois tipos de grelhas de discretização especialmente úteis.

Definição 3.7: Grelha de Discretização Padrão. Diz-se que G é uma Grelha de Discretização Padrão se for composta por um conjunto de células C_{ij} definidas por:

$$C_{ij} = [i, i + 1] \times [j, j + 1]; \quad i = 0, \dots, M - 1; \quad j = 0, \dots, N - 1$$

De acordo com esta definição, uma grelha de discretização padrão é uma grelha uniforme com células rectangulares de dimensões unitárias contíguas no espaço.

Definição 3.8: Grelha de Discretização Isomorfa. Seja $G^P = \{C_{ij}^P\}$ uma grelha de discretização padrão. Diz-se que $G^S = \{C_{ij}^S\}$ é uma grelha de discretização isomorfa a G^P sse existir uma transformação de coordenadas \mathbf{S} tal que:

$$C_{ij}^S = \mathbf{S}(C_{ij}^P)$$

e denota-se $G^S = \mathbf{S}(G^P)$.

O próximo teorema relaciona os dois métodos de aquisição de imagens de resolução não uniforme, referidos anteriormente, através da relação entre a transformação de coordenadas utilizada na criação da grelha de discretização no *mapeamento discreto* com a transformação de coordenadas aplicada no *mapeamento contínuo*.

Teorema 3.2: Discretizações equivalentes. Sejam:

- $I(x, y)$ uma imagem contínua;
- $I'(w, z)$ a imagem obtida a partir de $I(x, y)$ pela transformação de coordenadas $(w, z) = T(x, y)$;
- $G^P = \{C_{ij}^P\}$ uma grelha de discretização padrão;
- $G^S = \{C_{ij}^S\}$ uma grelha de discretização mapeada.

Consideremos os seguintes métodos de discretização:

1. discretização de $I'(w, z)$ pela grelha G^P , resultando em $I_1^d(i, j)$ (*mapeamento contínuo*);
2. discretização de $I(x, y)$ pela grelha G^S , resultando em $I_2^d(i, j)$ (*mapeamento discreto*).

Os dois processos são equivalentes se se verificarem as seguintes hipóteses:

1. a grelha G^S for obtida a partir de G^P pela transformação de coordenadas $\mathbf{S} = \mathbf{T}^{-1}$, ou seja, $G^S = \mathbf{T}^{-1}(G^P)$;
2. as células de G^S forem suficientemente pequenas tal que se possa considerar $I(x, y)$ constante em cada uma delas, ou seja, $I(x, y) = I_{ij}$, para $(x, y) \in C_{ij}^S$.

Demonstração:

As discretizações pelos métodos referidos resultam em:

$$I_1^d(i, j) = \frac{\iint_{C_{ij}^P} I'(w, z) \, dw dz}{\iint_{C_{ij}^P} \, dw dz}$$

e:

$$I_2^d(i, j) = \frac{\iint_{C_{ij}^S} I(x, y) \, dx dy}{\iint_{C_{ij}^S} \, dx dy}$$

Aplicando a mudança de coordenadas $(w, z) = \mathbf{T}(x, y)$ à expressão de $I_1^d(i, j)$, temos:

$$I_1^d(i, j) = \frac{\iint_{\mathbf{T}^{-1}(C_{ij}^P)} I(x, y) |\mathbf{J}(x, y)| \, dx dy}{\iint_{\mathbf{T}^{-1}(C_{ij}^P)} |\mathbf{J}(x, y)| \, dx dy}$$

Como, pela hipótese 1, $C_{ij}^S = \mathbf{T}^{-1}(C_{ij}^P)$, vem:

$$I_1^d(i, j) = \frac{\iint_{C_{ij}^S} I(x, y) |\mathbf{J}(x, y)| \, dx dy}{\iint_{C_{ij}^S} |\mathbf{J}(x, y)| \, dx dy}$$

Introduzindo a segunda hipótese, teremos:

$$I_1^d(i, j) = I_{ij} \frac{\iint_{C_{ij}^S} |\mathbf{J}(x, y)| \, dx dy}{\iint_{C_{ij}^S} |\mathbf{J}(x, y)| \, dx dy} = I_{ij}$$

e:

$$I_2^d(i, j) = I_{ij} \frac{\iint_{C_{ij}^S} \, dx dy}{\iint_{C_{ij}^S} \, dx dy} = I_{ij}$$

pele que, os dois métodos de discretização dados são equivalentes, q.e.d.

A segunda condição expressa no teorema anterior é muito restritiva. As imagens obtidas por um sistema robótico que interage no mundo real são bastante diversificadas o que significa que, em geral, os dois métodos referidos não são equivalentes.

Utilizando uma forma de discretização de imagem dada por:

$$I^d(i, j) = \frac{\int_{C_{ij}} I(x, y) |J(x, y)| dx dy}{\int_{C_{ij}} |J(x, y)| dx dy}$$

pode mostrar-se que os dois métodos tornam-se equivalentes. No entanto, não é claro qual dos métodos produz melhores resultados, pelo não se analisa em mais detalhe esta forma de discretização.

3.1.5 Imagens reamostradas.

Imagens reamostradas são imagens discretas obtidas por meio de outras imagens discretas com diferente resolução.

Muitos sistemas que utilizam imagens de resolução não uniforme fazem reamostragem de imagens discretas cartesianas, ou amostragem através de sensores com células não uniformes. Este facto deve-se à maior facilidade em obter imagens contínuas cartesianas do que imagens contínuas de resoluções arbitrárias, uma vez que esta última forma requer processamento de imagem no domínio contínuo, por exemplo através da utilização de lentes específicas.

As duas próximas definições indicam uma metodologia para obter imagens reamostradas.

Definição 3.9: Reconstrução de uma imagem contínua a partir de uma imagem discreta e de uma grelha de discretização. Seja $I(x, y)$ uma imagem contínua e $I^d(i, j)$ a imagem resultante da discretização de I por uma grelha $G = \{C_{ij}\}$.

Denota-se $I'(x, y)$ à reconstrução de $I(x, y)$ a partir de $I^d(i, j)$ e da grelha G , obtida por:

$$I'(x, y) = \begin{cases} I^d(i, j) & \Leftarrow (x, y) \in C_{ij} \\ 0 & \Leftarrow (x, y) \notin C_{ij} \end{cases}$$

Na generalidade dos casos, a discretização de imagem por este processo é um processo irreversível, ou seja, a reconstrução de uma imagem contínua feita pelo método indicado

na definição anterior, não é igual à imagem contínua original. De notar que este método de reconstrução é substancialmente diferente do método de reconstrução de sinais de banda limitada proveniente do teorema da amostragem [35]

Definição 3.10: Reamostragem. Seja $I^d(i, j)$ uma imagem resultante da discretização de uma imagem contínua $I(x, y)$ através da grelha G_1 .

Define-se como **reamostragem** de $I^d(i, j)$ pela grelha G_2 , como o processo de reconstruir $I(x, y)$ a partir de G_1 , resultando em $I'(x, y)$ e de seguida discretizar $I'(x, y)$ pela grelha G_2 , resultando $I^r(i, j)$. Às imagens $I^d(i, j)$ e $I^r(i, j)$ chamam-se **imagens semelhantes**.

Relembre-se que, no caso contínuo, as transformações de coordenadas permitiam-nos obter imagens isomorfas, às quais está associada uma bijecção, ou seja, existe correspondência ponto a ponto entre as duas imagens. O processo de reamostragem permite-nos criar imagens semelhantes mas, no caso geral, não se pode definir uma correspondência ponto a ponto: a reamostragem de imagens discretas é um processo irreversível, a não ser em casos particulares.

O procedimento mais comum para a obtenção de imagens discretas não uniformes, consiste na reamostragem, por uma grelha não uniforme, de imagens obtidas por um sensor de imagem uniforme. Cada célula constituinte da grelha de discretização uniforme do sensor de imagem será, daqui para a frente, denominada por *pixel*². Considerando o *pixel* como unidade de área, a resolução em cada ponto de uma imagem obtida por reamostragem será dada pelo recíproco do número de *pixels* existentes em cada célula da grelha não uniforme, como se encontra representado na Figura 3.5.

3.2 Transformação log-polar

No início deste capítulo foi referido que a integração de imagens com resolução não uniforme em sistemas de visão artificiais pode trazer vantagens no desempenho destes sistemas relativamente à utilização das habituais imagens cartesianas. No entanto existe

²do inglês, *picture element*.

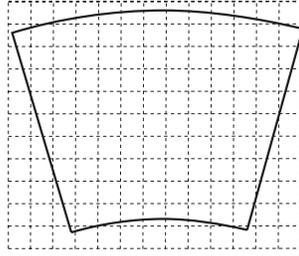


Fig. 3.5: A resolução de uma imagem reamostrada relaciona-se com o inverso do número de *pixel's* subjacentes a cada célula.

uma infinidade de transformações de coordenadas que permitem gerar imagens de resolução não uniforme. Portanto, a questão que se coloca é determinar a transformação a utilizar de modo a obter o máximo de vantagens.

A transformação log-polar tem sido a resposta preferida a esta questão por muitos investigadores [32, 63]. Por um lado argumenta-se a semelhança de tal transformação com a topologia da retina humana — em [57] refere-se que a distribuição dos fotoreceptores na retina segue uma lei do tipo logarítmica-polar — por outro lado demonstram-se propriedades desejáveis em sistemas de visão robótica tais como a invariância a escalamentos e/ou rotações [63], melhor fusão binocular [65], fácil computação do *tempo para contacto* [60] e compressão de imagem [27]. Por estes motivos iremos analisar em maior detalhe a transformação log-polar.

A transformação log-polar é descrita matematicamente por:

$$(\xi, \eta) = \mathbf{T}(x, y) = \left(\log_k \frac{\sqrt{x^2 + y^2}}{\rho_{\min}}, \alpha \arctan \frac{y}{x} \right); \quad (x, y) \in R \setminus (0, 0)$$

ou, numa notação mais compacta:

$$(\xi, \eta) = \left(\log_k \frac{\rho}{\rho_{\min}}, \alpha \theta \right)$$

em que (ρ, θ) representam as habituais coordenadas polares, tais como definidas no Exemplo 3.1. Os parâmetros ρ_{\min} , k e α são introduzidos para facilitar a definição das grelhas de discretização, como veremos adiante. Muitas vezes denominaremos ξ como coordenada radial e η como coordenada angular. Às imagens obtidas por esta transformação de coordenadas designaremos por **imagens log-polar**.

É fácil demonstrar as propriedades de invariância a rotações e escalamentos, em imagens log-polar. A invariância a rotações em torno da origem das coordenadas, ou invariância a escalamentos, não são propriedade exclusiva da transformação log-polar, quando consideradas individualmente. De facto, qualquer transformação que se possa exprimir por:

$$(w, z) = (f(\rho), \alpha \theta)$$

i.e. qualquer transformação rotacionalmente simétrica, exhibe invariância a rotações. Basta notar que após uma rotação de θ_0 , as novas coordenadas da imagem são dadas por:

$$(w_0, z_0) = (f(\rho), \alpha \theta + \alpha \theta_0)$$

o que corresponde a uma translacção na coordenada z , não havendo portanto alteração da forma. A invariância a escalamentos, por seu lado, advém da compressão logarítmica aplicada à coordenada radial. Aplicando um escalamento uniforme, de um factor s_0 , a uma transformação do tipo:

$$(w, z) = \left(\log_k \frac{\rho}{\rho_{\min}}, g(\theta) \right)$$

teremos:

$$(w_0, z_0) = \left(\log_k \frac{\rho}{\rho_{\min}} + \log_k s_0, g(\theta) \right)$$

o que corresponde a uma translacção simples na coordenada w , preservando a forma da imagem. A transformação log-polar integra a compressão radial e simetria rotacional, pelo que exhibe as duas propriedades em simultâneo.

Como é estabelecido pelo Teorema 3.1, a resolução de imagens log-polar, está directamente relacionada com o *Jacobiano* da transformação. Efectuando os cálculos necessários (ver Apêndice A) obtém-se:

$$|\mathbf{J}(x, y)| = \frac{\alpha}{\log k \cdot (x^2 + y^2)}$$

o que, expresso em coordenadas polares para comparação com a transformação polar (Exemplo 3.1), vem dado por:

$$|\mathbf{J}| = \frac{\alpha}{\log k \cdot \rho^2}$$

Observa-se um decaimento de resolução na coordenada radial mais acentuado do que para a transformação polar. Enquanto a resolução desta última evolui com o recíproco da distância à origem, para a transformação log-polar a resolução decai com o inverso do quadrado dessa distância. As curvas de resolução constante no domínio cartesiano, apresentadas na Figura 3.6, são circunferências concêntricas à semelhança da transformação polar. No entanto nota-se que existe um maior incremento no espaçamento entre circunferências sucessivas do centro para a periferia da imagem, do que para a transformação polar (ver Figura 3.2), revelando uma mais rápida perda de resolução.

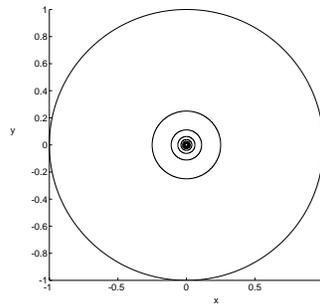


Fig. 3.6: Curvas de resolução constante para a transformação log-polar, representadas no plano cartesiano.

Uma das vantagens da utilização de imagens log-polar em sistemas de visão por computador advém da redução da quantidade de informação a processar. Entendida como a redução de área provocada pela aplicação de uma transformação de coordenadas entre zonas correspondentes nas duas imagens, então a redução da quantidade de informação é a grandeza recíproca da resolução. Desta forma, existirá redução de informação nas zonas do plano onde o Jacobiano da transformação de coordenadas for inferior à unidade.

No caso da transformação log-polar, o valor do Jacobiano toma valores muito altos em zonas próximas da origem, o que significa que é necessária uma grande quantidade de informação para as representar. A solução óbvia consiste em não representar a imagem numa zona próxima da origem, mantendo um compromisso entre o que se perde pelo facto de não representar uma parte da imagem e o que se ganha em redução da quantidade de informação na restante parte da imagem. Este assunto será um pouco mais aprofundado na próxima secção.

Para ilustrar o efeito da transformação log-polar sobre imagens reais, apresenta-se, na Figura 3.7, uma imagem cartesiana e o correspondente mapeamento para as coordenadas log-polares. A imagem log-polar foi obtida por reamostragem da imagem cartesiana por uma grelha log-polar (a criação destas grelhas será explicada em detalhe na próxima secção). O factor de redução de informação é de 1:8, neste caso (4K *pixels* na imagem log-polar para 32K *pixels* na imagem cartesiana).

As zonas mais à esquerda e mais à direita da imagem log-polar representam, respectivamente, a zona central e a zona periférica da imagem cartesiana. A zona periférica contém menos informação por unidade de área, por isso apresenta uma menor resolução. Este facto só se torna óbvio quando se procede à reconstrução da imagem para coordenadas cartesianas, apresentada na Figura 3.8. Como se pode observar, a zona central é visualizada com maior detalhe do que a zona periférica.

O ponto escuro existente no centro da imagem reconstruída deve-se ao facto da resolução tender para infinito na proximidade da origem. Embora a transformação log-polar não esteja definida na origem do plano cartesiano, a coordenada ξ toma valores arbitrariamente grandes (negativos) quando ρ tende para zero. Logo o tamanho das células de discretização necessárias para cobrir esta zona seria demasiado pequeno para o tamanho dos *pixels* subjacentes na imagem cartesiana, interrompendo-se por isso a discretização nesta área.



Fig. 3.7: Esquerda: Imagem cartesiana. Direita: Mapeamento para coordenadas log-polares.

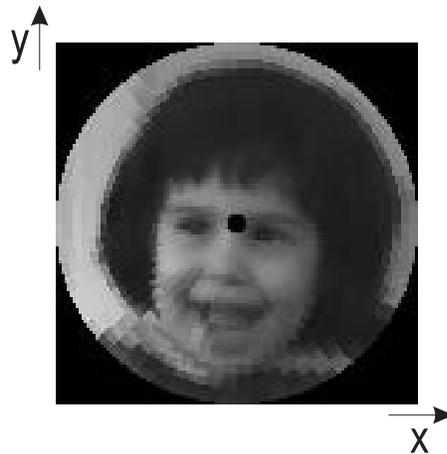


Fig. 3.8: Imagem reconstruída.

As imagens utilizadas no decorrer do trabalho experimental são obtidas por intermédio de mapeamento log-polar. Na próxima secção abordam-se, com algum detalhe, os aspectos relacionados com a sua construção e mecanismo de amostragem.

3.3 Emulação da retina

A aplicação de algoritmos de visão em computadores digitais requer a discretização de imagens contínuas. A discretização habitual de imagens é feita à custa de uma grelha de discretização uniforme aplicada a uma imagem contínua cartesiana, geralmente implementada através de dispositivos do tipo CCD.

Neste trabalho opta-se pela reamostragem por motivos de custo e flexibilidade. A utilização de dispositivos ópticos de transformação de coordenadas no domínio contínuo [16] ou a utilização de sensores de imagem com células dispostas de acordo com a geometria log-polar [51], para além de serem dispendiosos não permitem a redefinição fácil de parâmetros de amostragem, tais como o número de células de discretização ou a dimensão do campo visual.

A questão que se coloca resume-se à definição de uma grelha de discretização adequada a aplicar a imagens discretas cartesianas. A grelha a utilizar é obtida, por transformação inversa (ver Expressão (A.4)), a partir de uma grelha de discretização padrão. Desta

forma obtêm-se a denominada grelha de discretização log-polar, representada por:

$$C_{ij}^S = \left\{ \left(\rho_{\min} k^\xi \cos \frac{\eta}{\alpha}, \rho_{\min} k^\xi \sin \frac{\eta}{\alpha} \right) : (\xi, \eta) \in [i, i+1] \times [j, j+1] \right\};$$

com

$$i = 0, \dots, M-1; \quad j = 0, \dots, N-1$$

Na Figura 3.9 pode-se observar uma grelha de discretização log-polar obtida com dez passos de discretização em cada coordenada ($M = 10, N = 10$), e a grelha uniforme correspondente.

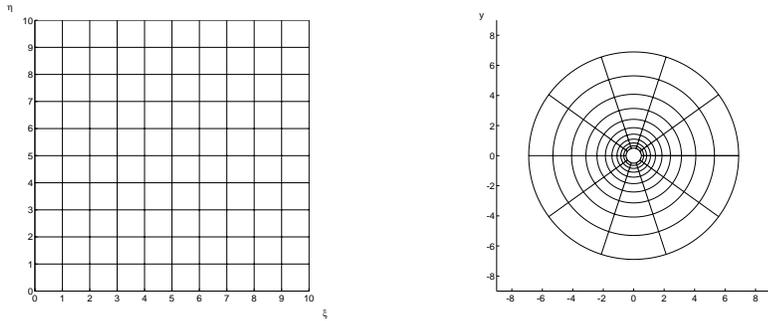


Fig. 3.9: Uma discretização uniforme do plano (ξ, η) corresponde a uma grelha log-polar no plano cartesiano.

3.3.1 Parâmetros da transformação

Os parâmetros α , k , e ρ_{\min} , são introduzidos de modo a facilitar a construção da grelha log-polar, em função da área coberta na imagem cartesiana e do número de passos de discretização utilizados.

O parâmetro α depende do número de intervalos de discretização pretendidos na coordenada angular (N) e é dimensionado para cobrir todo o espaço angular da imagem de entrada. Assim, $\eta = 0$ corresponde ao ângulo 0 rad. e $\eta = N$ corresponde ao ângulo 2π rad. pelo que:

$$\alpha = \frac{N}{2\pi}$$

A transformação log-polar apresenta uma singularidade na origem do plano. Desta forma, existe uma zona próxima da origem para a qual não se define a grelha de transformação. O parâmetro ρ_{\min} representa o raio desta vizinhança em torno da origem, que

não é representada em coordenadas log-polares. De facto, como $\rho = \rho_{\min} k^\xi$, temos:

$$\xi = 0 \Rightarrow \rho = \rho_{\min} k^0 = \rho_{\min}$$

Numa vizinhança da origem é necessária uma maior quantidade de informação para a representar do que a existente na imagem original. Uma forma simples de evitar a representação de informação redundante consiste em utilizar os *pixels* cartesianos subjacentes a essa zona e representá-los sem mapeamento. Outras formas mais elaboradas, que incluem a reamostragem por transformação hexagonal-polar e por vizinhanças de Voronoi, podem ser consultadas em [66]. Para o trabalho aqui apresentado, optou-se simplesmente por não representar uma zona próxima da origem tal que não provocasse um número demasiado alto de células de discretização e fosse bastante mais pequena do que os objectos envolvidos nas experiências.

Para definir o parâmetro k , admite-se que M é o número de passos de discretização na direcção radial e que queremos cobrir um campo de visão com um raio ρ_{\max} . Assim:

$$\xi = M \Rightarrow \rho = \rho_{\max} = \rho_{\min} k^M$$

o que define o valor do parâmetro k :

$$k = \left(\frac{\rho_{\max}}{\rho_{\min}} \right)^{1/M}$$

Resumindo, os parâmetros da transformação log-polar α , k e ρ_{\min} , estão directamente relacionados com o número de células de amostragem e com o campo de visão pretendido. Definindo estes valores (M , N e ρ_{\max}) fica completamente estabelecida a transformação a aplicar às imagens. Resta considerar os seguintes pontos:

1. O valor de ρ_{\max} não poderá ser tal que exceda as dimensões da imagem cartesiana — muitas vezes opta-se por cobrir o máximo campo de visão possível.
2. O valor de ρ_{\min} não deverá ser demasiado grande, tal que uma grande parte da imagem não seja representada, nem deverá ser demasiado pequeno, tal que provoque um número muito grande de passos de discretização na coordenada radial.

O efeito da alteração dos parâmetros de projecto na forma das grelhas de amostragem log-polar, pode ser observado nas próximas figuras. Incluem-se esquemas com a geometria dos sensores e tabelas com diversos parâmetros de interesse: os parâmetros de projecto M , N e ρ_{\max} , os parâmetros da transformação α , k e ρ_{\min} e o índice de compressão de imagem.

O primeiro exemplo (Figuras 3.10 e 3.11) ilustra o aumento do número de passos de discretização, quer na direcção radial quer na direcção angular, para uma mesma área coberta na imagem cartesiana. O segundo caso (Figura 3.12) mostra o efeito do aumento do raio mínimo, para igual campo de visão, como se fossem removidas células da parte interior da grelha. Por fim ilustram-se dois casos em que a dimensão das imagens log-polar não se altera, mas dão origem a grelhas cobrindo diferentes campos visuais:

- O escalamento uniforme do campo visual (Figura 3.13) em que a dimensão radial das células é multiplicada por um factor constante.
- O escalamento não uniforme do campo visual (Figura 3.14) em que a taxa de crescimento da dimensão radial das células, aumenta de um sensor para o outro (variação de k).

A escolha de uma geometria específica para um sensor log-polar deve ter em conta factores tais como: a dimensão da imagem original, a dimensão desejada para as imagens mapeadas (que pode depender da capacidade de processamento disponível) e o compromisso pretendido entre a resolução angular e resolução radial. A influência que cada conjunto específico de parâmetros tem no desempenho dos algoritmos de processamento utilizados terá que ser analisada para cada caso. Por exemplo, em [13] são apresentados resultados da influência dos parâmetros da transformação log-polar no desempenho de algoritmos de correlação em processos de vergência.

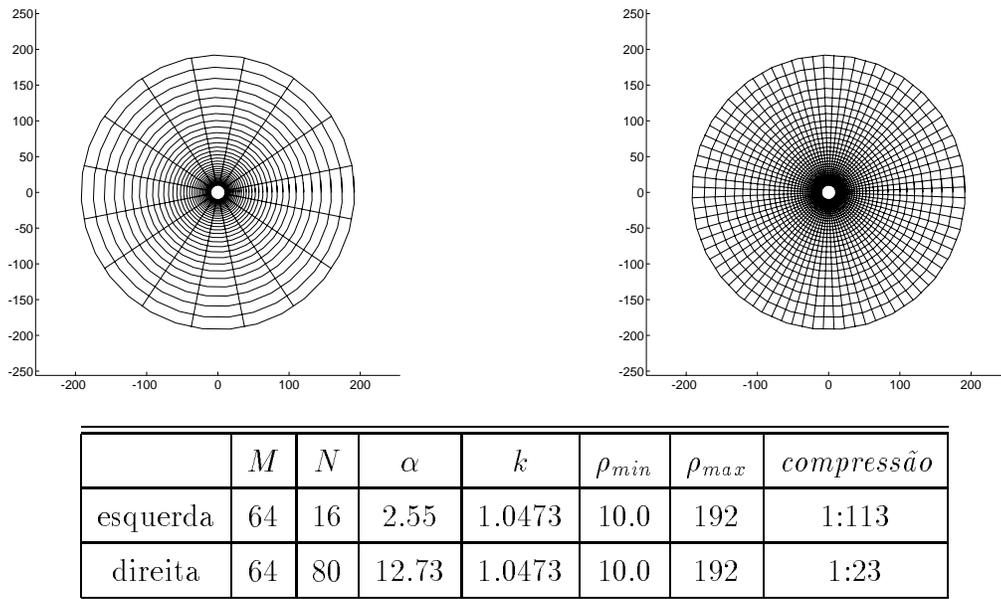


Fig. 3.10: Grelhas de discretização log-polar para diferentes resoluções angulares.

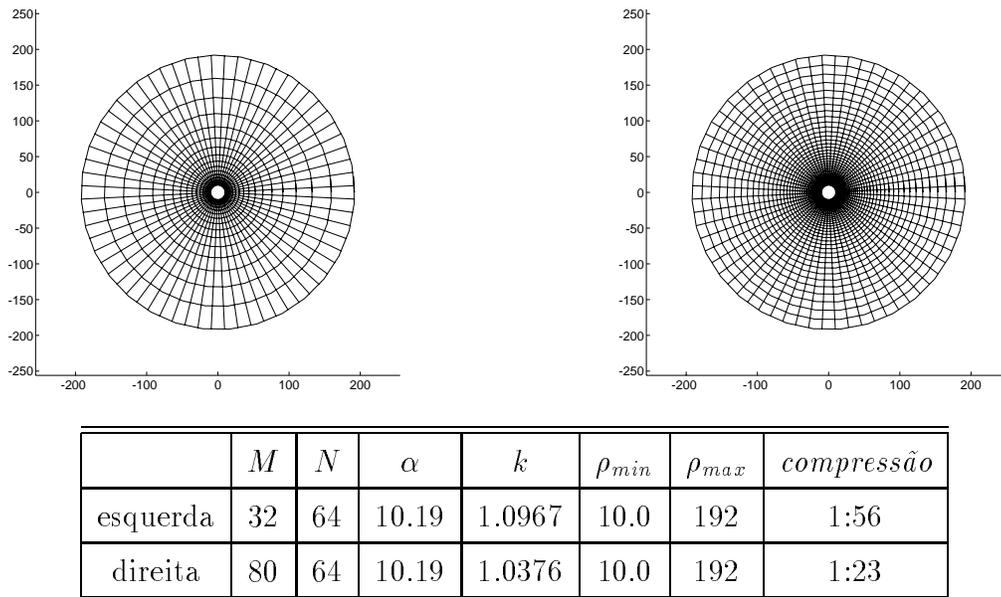


Fig. 3.11: Grelhas de discretização para diferentes resoluções radiais.

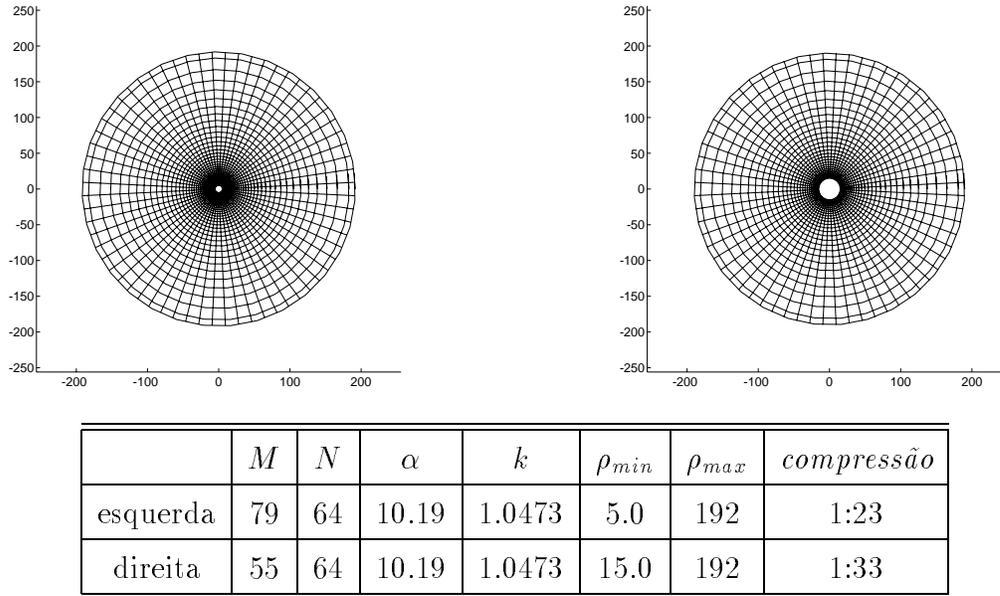


Fig. 3.12: Grelhas de discretização log-polar para diferentes valores do raio mínimo

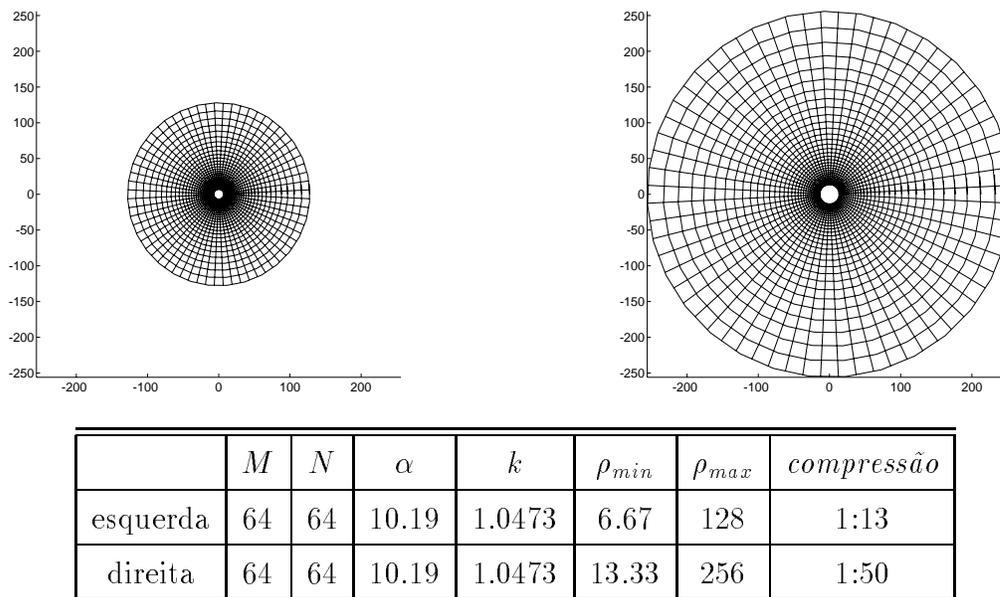


Fig. 3.13: Grelhas de discretização log-polar sob o efeito de um escalamento uniforme do campo visual.

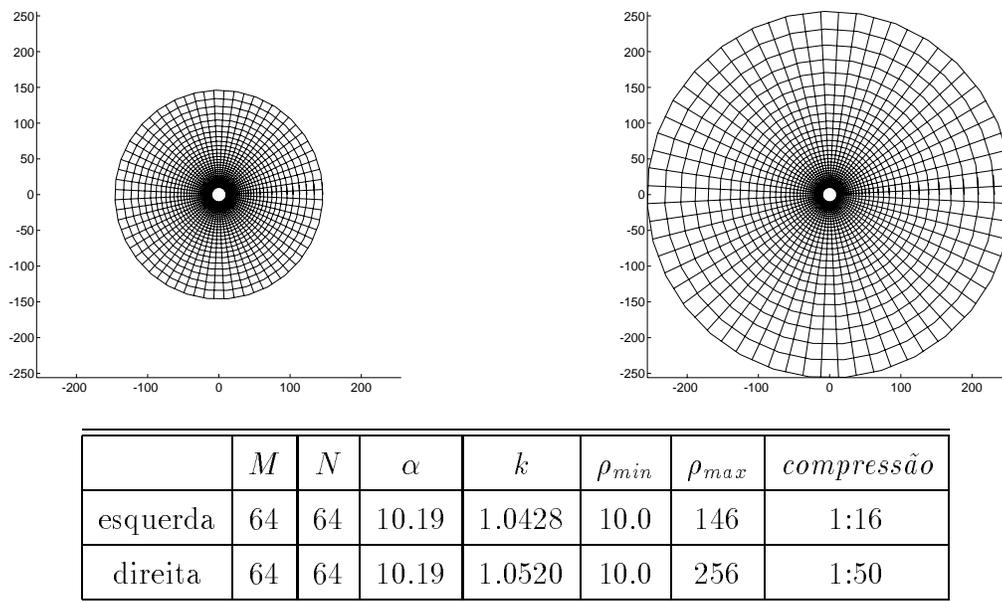


Fig. 3.14: Grelhas de discretização log-polar sob o efeito de um escalamento não uniforme do campo visual.

4. Vergência

A vergência é um dos movimentos básicos do sistema visual humano [21]. Tem como objectivo principal manter o alvo de interesse na zona central das retinas quando este se desloca em profundidade, sendo constituído por rotações oculares em sentidos contrários.

Devido ao posicionamento dos olhos no ser humano, os movimentos de vergência horizontal são, de longe, os mais importantes. Sob condições artificiais podem ser produzidos movimentos de vergência vertical [21] mas estes movimentos são muito limitados em amplitude e não ocorrem naturalmente. O modo de quantificar o grau de vergência numa determinada situação recorre à medida do ângulo formado pelas duas linhas de observação (**ângulo de vergência**), como representado na Figura 4.1.

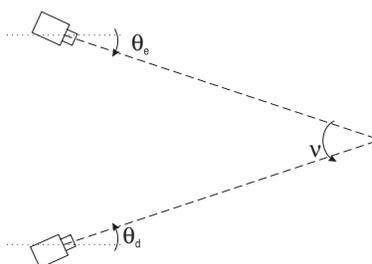


Fig. 4.1: Duas câmaras vergem num ponto. O ângulo de vergência é representado por ν .

Os movimentos de vergência são fundamentalmente involuntários: embora muitos indivíduos consigam vergir ou divergir voluntariamente sem informação visual, normalmente os movimentos de vergência são executados em resposta a estímulos visuais provocados por objectos que se deslocam em profundidade. Por outro lado, experiências com humanos mostram que a apresentação de imagens diferentes em cada olho provoca uma resposta rápida do sistema de vergência [21]. Estes dois factos sugerem a existência de mecanismos de controlo de vergência com reduzida intervenção cognitiva e que utilizam informação

visual de baixo nível, numa fase prévia ao reconhecimento dos objectos existentes na cena.

4.1 Estímulos para controlo de vergência

O ângulo de vergência está directamente relacionado com a distância do alvo ao observador. Os indícios visuais de profundidade constituem, por isso, importantes sinais de entrada para o sistema de controlo de vergência. Entre os mais influentes encontram-se a disparidade binocular e a desfocagem retinal, embora indícios visuais tais como perspectiva, dimensão e sobreposição de objectos, sejam utilizados pelos humanos. Estes últimos, por serem influenciados por factores cognitivos, são classificados como indícios “psíquicos” ou de alto nível [21].

Os estímulos de disparidade e de acomodação podem cooperar no processo de vergência [37]. O estímulo de disparidade binocular surge quando o objecto de interesse tem localizações diferentes nas duas retinas. Nos humanos, este estímulo, por si só, induz movimentos de vergência compensatórios de modo a reduzir o grau de disparidade. Quando este valor é muito baixo diz-se que existe fusão binocular. O estímulo de acomodação é composto basicamente pela desfocagem da imagem presente na retina. Como resposta a este estímulo observa-se uma alteração da curvatura ou espessura da lente ocular de modo a focar a imagem do objecto de interesse.

O sistema visual humano apresenta dois tipos de acoplamento entre os movimentos de vergência e de acomodação:

- a vergência por acomodação — quando um dos olhos foca um objecto por acomodação, então os eixos ópticos são orientados de modo a vergir nesse objecto;
- a acomodação por vergência — o processo de vergir num objecto, como resposta a um estímulo de disparidade, provoca a acomodação dos olhos de forma a focar para a distância de vergência.

A distâncias longas, quando a sensibilidade da desfocagem relativamente a alterações de profundidade é muito pequena, o estímulo de disparidade é dominante. Por outro lado, para distâncias curtas, o processo de fusão binocular é menos fiável devido à deformação perspectiva, pelo que a informação de desfocagem se torna mais útil.

Neste capítulo iremos discutir apenas a aplicação de estímulos de disparidade retinal na implementação de um sistema de vergência. Isto deve-se, não só pelo facto de se utilizarem lentes de focagem manual, mas também porque a disparidade constitui um estímulo mais poderoso para uma grande gama de profundidades. No entanto, a utilização de estímulos de desfocagem pode ser benéfica, principalmente em zonas muito próximas do observador.

4.2 Correlação para controlo de vergência

Consideremos uma configuração de fixação frontal de um alvo. Se o alvo se mover ao longo do eixo ciclópico (ver Figura 4.2) vai provocar uma disparidade retinal não nula. O problema de controlo de vergência consiste em tentar anular esta disparidade, garantindo uma nova fixação do alvo.

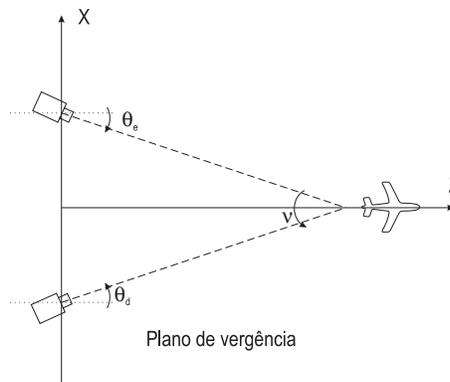


Fig. 4.2: Objectos presentes no eixo ciclópico não produzem disparidades horizontais nulas, excepto se coincidirem com o ponto de fixação.

A disparidade é uma das medidas preferencialmente utilizadas para o controlo da vergência em sistemas de visão artificiais; uma vez obtido o valor de disparidade é possível, num só passo, controlar os ângulos das câmaras de modo a atingir a vergência correcta.

Os algoritmos habitualmente utilizados para calcular medidas explícitas de disparidade incluem o filtro de *cepstrum* [40], métodos de análise de fase [29, 67] e procura por correlação (*template matching*) [35], entre outros, mas todos eles revelam-se computacionalmente pesados, e conseqüentemente, pouco indicados para aplicações em tempo-real. Nesta secção analisam-se alternativas a esses métodos, baseadas em processos mais

simples e rápidos e que permitem atingir situações de vergência correcta mesmo com medidas de menor precisão.

4.2.1 Medidas de correlação entre imagens

A aplicação de métodos de correlação a duas imagens, proporciona a obtenção de estimativas da semelhança entre as imagens. Permitem portanto, a detecção de situações de vergência, onde o alvo ocupa a mesma posição nas duas retinas, provocando uma elevada correlação. Para além disso, são métodos computacionalmente rápidos e matematicamente simples, permitindo simultaneamente um bom desempenho em tempo-real e uma fácil análise do comportamento dos algoritmos. Por estes factores, iremos utilizar métodos de correlação de imagem para obter a informação necessária para o controlo da vergência.

Existem muitos algoritmos de correlação documentados na literatura específica de áreas como a fotogrametria e a visão *stereo*. Em [7] encontra-se um estudo comparativo de vários algoritmos conhecidos de correlação, examinados em diversas condições, incluindo alterações de brilho médio entre as imagens, adição de ruído e distorções controladas. Uma das medidas de correlação mais conhecidas e simples é obtida por acumulação espacial do quadrado da diferença entre as imagens e é habitualmente denominada por *SSD* (*Sum of Squared Differences*), na sua versão discreta. Sendo $I_1(i, j)$ e $I_2(i, j)$ duas imagens discretas, o valor da *SSD* é dado por:

$$SSD(I_1, I_2) = \sum_{i,j} [I_1(i, j) - I_2(i, j)]^2$$

A *SSD* é, por razões históricas e simplicidade, uma das medidas de correlação mais utilizadas. Advém da interpretação discreta da distância \mathcal{D} entre duas imagens contínuas no espaço das funções escalares de variável bidimensional, segundo a norma usual:

$$\mathcal{D}(I_1, I_2) = \|I_1 - I_2\|_2 = \iint_{x,y} (I_1(x, y) - I_2(x, y))^2 dx dy \quad (4.1)$$

Dado que a *SSD* é habitualmente incluída na classe das medidas de correlação, o resultado da sua aplicação a duas imagens será denominado como o índice de correlação entre essas imagens. Assim, para esta medida, a semelhança entre duas imagens será máxima quando o índice de correlação for mínimo.

4.2.2 Correlação em imagens de resolução não uniforme

No estudo da aplicação de geometrias de resolução não uniforme na representação das imagens, interessa-nos avaliar a influência da aplicação dessas geometrias sobre o desempenho dos algoritmos de correlação e, por consequência, no processo de vergência. Apesar de não ser aplicada computacionalmente, a versão contínua da SSD, dada pela Equação (4.1), permite analisar mais facilmente essa influência, uma vez que as imagens de resolução não uniforme são geradas pela aplicação de transformações de coordenadas que se encontram definidas em domínios contínuos. O próximo teorema estabelece uma importante relação entre a correlação de imagens de resolução variante no espaço e a correlação de imagens cartesianas.

Teorema 4.1: Correlação de imagens de resolução não uniforme Sejam duas imagens, $I'_1(w, z)$ e $I'_2(w, z)$, obtidas a partir das imagens cartesianas $I_1(x, y)$ e $I_2(x, y)$, pela transformação de coordenadas $(w, z) = \mathbf{T}(x, y)$. Seja ainda $|\mathbf{J}|$ o jacobiano da transformação de coordenadas \mathbf{T} , e \mathcal{D} o operador de distância entre imagens, segundo a Equação (4.1). Nestas condições, verifica-se a seguinte igualdade:

$$\mathcal{D}(I'_1, I'_2) = \mathcal{D}\left(I_1\sqrt{|\mathbf{J}|}, I_2\sqrt{|\mathbf{J}|}\right)$$

Demonstração:

$$\mathcal{D}(I'_1, I'_2) = \iint_{w,z} (I'_1(w, z) - I'_2(w, z))^2 dw dz$$

Aplicando a mudança de coordenadas $(w, z) = \mathbf{T}(x, y)$, vem:

$$\mathcal{D}(I'_1, I'_2) = \iint_{x,y} (I_1(x, y) - I_2(x, y))^2 |\mathbf{J}(x, y)| dx dy$$

o que equivale a:

$$\mathcal{D}(I'_1, I'_2) = \iint_{x,y} \left(I_1(x, y) \sqrt{|\mathbf{J}(x, y)|} - I_2(x, y) \sqrt{|\mathbf{J}(x, y)|} \right)^2 dx dy$$

e, portanto:

$$\mathcal{D}(I'_1, I'_2) = \mathcal{D}\left(I_1\sqrt{|J|}, I_2\sqrt{|J|}\right)$$

q.e.d.

Assim, o índice de correlação entre duas imagens mapeadas equivale ao índice de correlação calculado sobre as imagens cartesianas originais, ponderadas pela raiz quadrada do jacobiano da transformação de coordenadas. Significa isto que as características das imagens cartesianas, em locais onde o jacobiano é elevado, têm uma maior influência no cálculo do índice de correlação.

O seguinte corolário particulariza o Teorema 4.1 para a transformação log-polar. Como vimos no Capítulo 3, o jacobiano da transformação log-polar é inversamente proporcional ao quadrado da distância à origem, em coordenadas cartesianas. Assim o cálculo do índice de correlação entre duas imagens log-polar corresponde ao cálculo do índice de correlação entre as imagens cartesianas isomorfas, afectadas de um factor de ponderação inversamente proporcional à distância à origem.

Corolário 4.2. – Correlação de imagens log-polar Sejam duas imagens, $I'_1(w, z)$ e $I'_2(w, z)$, obtidas a partir das imagens cartesianas $I_1(x, y)$ e $I_2(x, y)$, pela transformação de coordenadas log-polar. Seja \mathcal{D} o operador de distância entre imagens, segundo a equação (4.1). Nestas condições, verifica-se a seguinte igualdade:

$$\mathcal{D}(I'_1, I'_2) = \mathcal{D}\left(\frac{I_1(x, y)}{\beta\sqrt{x^2 + y^2}}, \frac{I_2(x, y)}{\beta\sqrt{x^2 + y^2}}\right)$$

sendo β um escalar dependente dos parâmetros da transformação.

Demonstração:

O Jacobiano da transformação log-polar é dado por:

$$|\mathbf{J}(x, y)| = \frac{\alpha}{\log k} \cdot \frac{1}{x^2 + y^2}$$

e a sua raiz quadrada:

$$\sqrt{|\mathbf{J}(x, y)|} = \frac{1}{\beta\sqrt{x^2 + y^2}}$$

com:

$$\beta = \sqrt{\frac{\log k}{\alpha}}$$

portanto, pelo Teorema 4.1, demonstra-se o resultado pretendido.

Segundo esta interpretação, podemos considerar a correlação de imagens log-polar como a correlação de imagens cartesianas multiplicadas por uma máscara (ou janela espacial) que pondera favoravelmente a zona central das imagens em detrimento das zonas periféricas. A Figura 4.3 apresenta uma representação gráfica deste factor de ponderação. A correlação de imagens log-polar, ao dar maior atenção à zona central do campo visual, implementa indirectamente um mecanismo de atenção dirigido para essa zona.

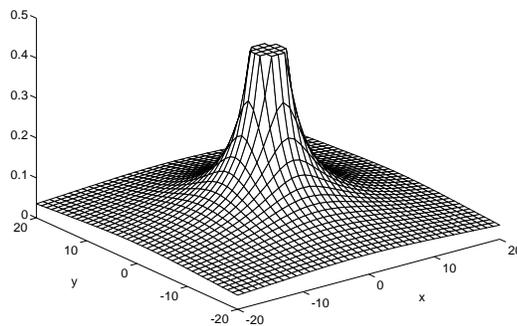


Fig. 4.3: Janela de ponderação associada à transformação log-polar. A zona no centro das coordenadas encontra-se truncada devido à singularidade da transformação.

A questão que agora se coloca reside em saber se este mecanismo atencional implícito é ou não benéfico para os objectivos em causa. Intuitivamente, para o controlo da vergência, esta característica traz importantes vantagens sobre a correlação simples de imagens cartesianas. Quando as câmaras vergem num objecto posicionado no centro do campo visual, então a área do objecto sob vergência tem disparidades baixas e a área na periferia da imagem, pertencente ao fundo, exhibe disparidades elevadas. Para imagens log-polar, os pontos da periferia têm uma menor influência no cálculo da correlação do que os pontos no centro da imagem o que resulta num valor alto de semelhança entre as imagens. Com imagens cartesianas, onde as disparidades são consideradas equitativamente ao longo de todo o domínio, as disparidades do fundo podem dar origem a valores altos no índice de correlação, prejudicando a detecção das situações de vergência.

Para ilustrar a vantagem da utilização da geometria log-polar sobre a geometria carte-

siana na detecção de configurações correctas de vergência, foi efectuada uma experiência simples. Colocou-se um objecto estático em frente das câmaras e calcularam-se os valores de correlação entre os pares de imagens obtidas, para um amplo intervalo de ângulos de vergência. Os valores obtidos podem ser considerados como a amostragem de uma função contínua $f(\nu)$ que corresponde à correlação entre as imagens com um ângulo de vergência ν , e que será denominada por **função de correlação**.

Na Figura 4.4 representam-se graficamente os resultados obtidos utilizando duas representações log-polar e uma cartesiana. Pode observar-se que as amostras obtidas com representações log-polar organizam-se segundo perfis com um mínimo absoluto bem evidente no ponto correspondente ao ângulo de vergência correcto, o que torna simples a sua detecção. Por outro lado, o perfil obtido com a representação cartesiana não distingue com clareza o mínimo absoluto dos mínimos locais, pelo que não se adequa ao objectivos pretendidos.

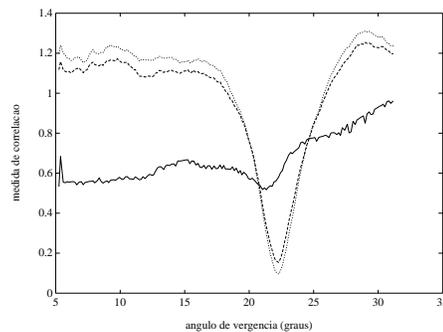


Fig. 4.4: Perfis de correlação como função do ângulo de vergência para uma situação real, com representações cartesianas (linha a cheio) e log-polar (linhas tracejada e pontilhada).

Foram efectuadas diversas experiências do tipo atrás descrito, em situações reais, e na maior parte delas, as configurações de vergência correcta corresponderam aos mínimos globais da função de correlação com imagens log-polar. No entanto, com imagens cartesianas correspondência não se verifica em muitas situações. De facto, se a área da imagem ocupada pelo alvo for inferior à área do fundo, então a disparidade do fundo torna-se dominante e a detecção de vergência falha.

Enquanto que com representações cartesianas é normal que a área do alvo seja pequena relativamente ao fundo, causando falhas na detecção de vergência, com representações log-

-polar conseguimos um melhor controlo dessas situações, por adequação dos parâmetros da transformação às dimensões dos objectos. Considera-se que a influência do alvo é dominante no cálculo da correlação se a área que ocupa na imagem log-polar for superior a metade da área total da imagem. Por exemplo, admitindo que o alvo a fixar se mantém no centro do campo visual e tem forma circular, então a percentagem mínima da área total cartesiana que deverá ocupar para que a sua influência seja dominante é designada por área crítica (A_c) e é dada por:

$$A_c = \frac{\rho_{\min}}{\rho_{\max} + \rho_{\min}} \times 100\%$$

Portanto, valores baixos da relação ρ_{\min}/ρ_{\max} conduzem a áreas críticas baixas, o que significa que mesmo objectos com dimensões pequenas poderão ter áreas superiores à área crítica. Como contrapartida, a quantidade de informação utilizada para representar a imagem aumenta à medida que se diminui a razão ρ_{\min}/ρ_{\max} . O projecto do sensor log-polar deverá ter em conta estes factos, adequando-se à dimensão dos objectos que se pretendem fixar e à gama de profundidades para as quais se pretende que o processo funcione.

4.3 Algoritmos de controlo de vergência

Partindo do pressuposto que, para uma determinada configuração estática do campo visual, existe uma função de correlação com um mínimo absoluto bem saliente, correspondendo ao ângulo de vergência correcto, pode-se encarar o problema do controlo de vergência como um problema de minimização: procurar o valor do ângulo de vergência que minimiza a função de correlação, ao longo de um intervalo suficientemente largo.

Nesta secção apresentam-se dois algoritmos de controlo do processo de vergência, utilizando técnicas de procura na função de correlação e controlo por retroacção. O controlo não é efectuado num só passo, uma vez que implicaria o cálculo preciso da disparidade e o conhecimento da calibração do sistema, mas sim de uma forma iterativa, através da integração das medidas de correlação obtidas ao longo do tempo, com uma estratégia de controlo em malha fechada. Esta estratégia permite a convergência para a solução desejada apesar da menor precisão dos cálculos efectuados e da falta da informação de

calibração sobre o sistema.

4.3.1 Controlo por optimização local

A forma mais directa de procurar o mínimo global da função de correlação numa determinada configuração visual implica a variação do ângulo de vergência ao longo de um intervalo de valores suficientemente amplo. Do ponto de vista prático esta estratégia não é muito adequada dado que obriga a movimentação das câmaras e o processamento das imagens adquiridas, para um número relativamente grande de pontos no espaço visual, o que dá origem a ritmos de amostragem demasiado baixos. Por outro lado, o ambiente tem que se manter estático durante o processo de recolha de informação, uma vez que cada função de correlação só é válida para uma determinada configuração estática do campo visual. Uma das alternativas a este tipo de optimização global consiste em efectuar uma optimização local partindo do ângulo de vergência corrente, através, por exemplo, de um método de gradiente. A convergência para a solução desejada supõe que esta esteja suficientemente próxima do ângulo de vergência inicial, ou corre-se o risco de convergir para um mínimo local. Há que admitir também que a cena se mantém estática durante o processo de procura.

A minimização por método de gradiente é uma das formas mais utilizadas na optimização de funcionais em espaços de procura desconhecidos. Baseia-se na adaptação dos parâmetros livres em função do gradiente do funcional de custo. Para um determinado ângulo de vergência (ν) e uma determinada situação visual caracterizada pela função de correlação $f(\nu)$, a adaptação do valor do ângulo de vergência é dada por:

$$\Delta\nu = -\eta\nabla_{\nu}f(\nu)$$

em que η é o passo de adaptação e $\nabla_{\nu}f(\nu)$ designa o gradiente da função de correlação.

No caso que estamos a considerar, o valor do gradiente não pode ser explicitamente calculado. A sua estimação recorre a medidas diferenciais da função de correlação e dos valores do ângulo de vergência:

$$\nabla_{\nu}f(\nu)^i \approx \frac{f(\nu^i) - f(\nu^{i-1})}{\nu^i - \nu^{i-1}}$$

onde o índice i corresponde ao instante em que as medidas são obtidas.

Este processo de estimação do valor do gradiente não se mostrou satisfatório devido ao nível de ruído existente nos sinais em questão e à quase singularidade existente na proximidade do mínimo da função de correlação (o termo $\nu^i - \nu^{i-1}$ toma valores próximos de zero). A abordagem adoptada para o controlo da vergência evita a utilização dos valores absolutos de medidas diferenciais utilizando apenas a informação sobre o seu sinal. O algoritmo utilizado segue a seguinte lei:

$$\Delta\nu^i = f(\nu^i) \cdot [-\text{sgn}(\Delta f(\nu^i))] \cdot \text{sgn}(\Delta\nu^{i-1})$$

onde $\Delta f(\nu^i) = f(\nu^i) - f(\nu^{i-1})$. Quando sucessivos ângulos de vergência correspondem a decréscimos de $f(\nu)$, ou seja, a procura decorre no sentido correcto, então a adaptação do ângulo preserva esse sentido. Por outro lado, se o sentido de procura conduz a valores crescentes de $f(\nu)$, então a adaptação do ângulo de vergência inverte esse sentido. O factor multiplicativo $f(\nu^i)$ provoca adaptações grandes quando o valor do ângulo de vergência difere muito da solução pretendida, ou adaptações pequenas quando nos encontramos nas proximidades da vergência correcta.

Nas muitas experiência efectuadas, este algoritmo demonstrou um bom desempenho no que respeita à convergência para a solução desejada. A sua sensibilidade ao ruído é bastante inferior à do método de gradiente simples, como se ilustra na Figura 4.5. No

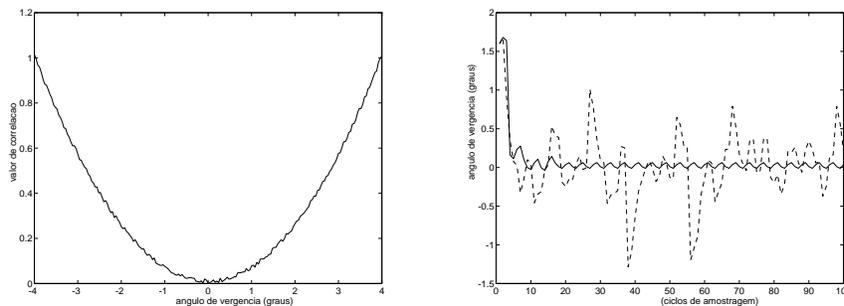


Fig. 4.5: À esquerda: Função de correlação parabólica com ruído aditivo. À direita: Simulação do sistema de vergência com método de gradiente simples (linha interrompida) e com o algoritmo proposto (linha contínua).

entanto, dois pontos negativos têm que ser notados:

1. as câmaras têm que estar permanentemente em movimento de modo a detectar as variações do valor da função de correlação necessárias para a lei de adaptação, o que provoca oscilações em torno do ângulo correcto de vergência;
2. o valor mínimo da função de correlação toma valores diferentes consoante as disparidades e áreas relativas entre o alvo e o fundo, pelo que o comportamento do processo de procura, no que respeita ao tempo de convergência e à amplitude das oscilações, será distinto de uma situação para outra.

O algoritmo que se descreve na próxima secção tenta solucionar estes problemas assim como relaxar as restrições impostas relativamente ao dinamismo do ambiente durante o processo de procura e à proximidade entre a condição inicial e a solução pretendida.

4.3.2 Controlo por optimização global

A optimização global por variação directa do ângulo de vergência num dado intervalo, tem inconvenientes numa implementação prática. O facto de requerer a movimentação mecânica das câmaras para efectuar o “varrimento” dos ângulos de vergência, torna o processo demasiado lento para a pretendida funcionalidade em tempo real. A solução que aqui se propõe tenta virtualizar este processo, obtendo os valores da função de correlação para determinados ângulos de vergência sem necessitar de movimentar as câmaras. O processo de controlo resumir-se-à a procurar o mínimo desses valores e atribuir às câmaras o ângulo de vergência correspondente.

Para uma determinada configuração do ambiente, na qual se adquire um par *stereo* de imagens, podem obter-se aproximações dos valores da função de correlação em ângulos vizinhos do ângulo corrente, por cálculo da correlação em imagens adequadamente trasladadas na direcção horizontal. A aproximação é boa para pequenas variações do ângulo de vergência embora para variações grandes, as deformações de perspectiva tornem a aproximação mais grosseira.

Assim, a solução passa por obter aproximações da função de correlação num conjunto discreto de ângulos de vergência e procurar o mínimo desses valores. A determinação do número de amostras a obter e a sua distribuição pelos ângulos de vergência são as

questões que agora se colocam. Para as solucionar procede-se à aquisição das amostras de um modo não uniforme, pelas razões que a seguir se descrevem:

- quando o ângulo de vergência corrente está próximo do valor desejado, deverão ser efectuados pequenos ajustes, pelo que a densidade de amostras será grande numa zona próxima do ângulo corrente;
- quando o ângulo de vergência corrente está longe do valor desejado, interessa fundamentalmente atraí-lo para uma zona próxima, pelo que a estimativa da localização do mínimo pode ser mais grosseira.

Utilizando estas ideias pode-se cobrir um intervalo relativamente grande de ângulos de vergência com um número baixo de amostras. A convergência para o valor correcto é feita, de um modo geral, em duas fases: inicialmente o ângulo de vergência é atraído para as proximidades do ângulo correcto, de uma forma pouco precisa, e numa segunda fase procede-se ao ajuste fino até se atingir a solução pretendida.

Este algoritmo não pressupõe o conhecimento da relação existente entre a “quantidade” de translacção horizontal aplicada às imagens e a variação de ângulo de vergência a que essa translacção corresponde. Permite, no entanto, obter medidas do sentido e da amplitude relativa das adaptações a provocar ao ângulo de vergência. A integração destas medidas num sistema em malha fechada possibilita a convergência para a solução desejada apesar da falta de conhecimento exacto sobre a cinemática do sistema.

Cada ponto “amostrado” da função de correlação é calculado através da correlação entre imagens previamente transformadas por uma translacção horizontal. A definição dos pontos de amostragem é efectuada não através da especificação das variações do ângulo de vergência a efectuar, mas sim através da especificação das translacções a aplicar às imagens.

Outra interpretação possível é a de associar cada ponto de amostragem a um **canal de disparidade**, sintonizado para um nível de translacção horizontal pré-definido. Interpretado desta forma, cada canal terá como saída uma medida da verosimilhança da disparidade real entre as imagens ser dada pela disparidade associada ao canal. O algoritmo pode ser representado pelo diagrama da Figura 4.6.

Podemos resumir o algoritmo proposto da seguinte forma:

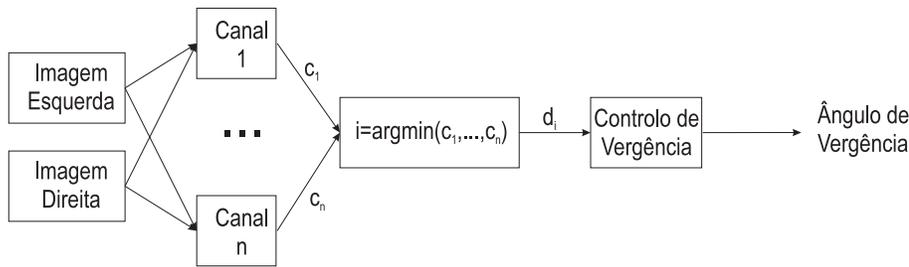


Fig. 4.6: Controlo de vergência por optimização global.

1. Consideremos um conjunto de n canais de disparidade, associado a um conjunto de disparidades horizontais $D = \{d_1, \dots, d_n\}$. Como já foi referido, disparidades elevadas devem ser seleccionadas de uma forma espaçada enquanto disparidades baixas devem cobrir densamente a zona associada.
2. A cada canal i é atribuído um mapa de correspondências entre os *pixels* das duas imagens, admitindo uma translação horizontal de d_i pixels de uma para a outra. Embora, para imagens cartesianas, uma translação horizontal seja um processo extremamente simples de efectuar, com imagens log-polar não acontece o mesmo. A Figura 4.7 ilustra a deformação das linhas horizontais quando mapeadas em coordenadas log-polar. Para cada valor d_i há que calcular o mapa de disparidades correspondente, no plano log-polar. Este passo é feito *a priori* e os resultados são guardados em tabelas (uma por canal).

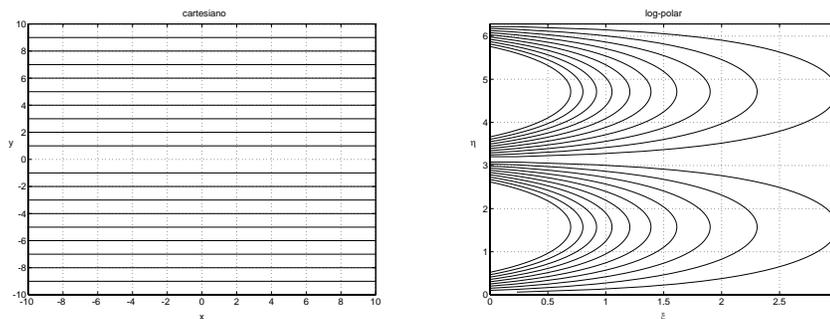


Fig. 4.7: Trajectórias de pontos sujeitos a uma translação horizontal, no plano cartesiano (esquerda) e no plano log-polar (direita)

3. Em cada instante, é obtido um par de imagens log-polar. Para cada canal i , as

imagens são transformadas de acordo com o mapa de disparidades correspondente e correlacionadas entre si, tendo como resultado o valor c_i .

- Os valores de saída de cada canal (c_i), são comparados entre si. Se j é o índice do canal com valor mínimo, então a disparidade d_j é usada como valor de referência a um processo de controlo em malha fechada. Esta disparidade pode ser considerada como a disparidade contida no conjunto D que melhor aproxima a disparidade dominante nas imagens.

Experimentalmente obtiveram-se os valores de saída dos canais para pares de imagens com disparidades conhecidas. Os resultados obtidos com imagens cartesianas e log-polar encontram-se representados na Figura 4.8, observando-se que, em geral, o canal com valor mínimo corresponde à melhor aproximação para a disparidade presente nas imagens. Em todo o caso, para as imagens log-polar, o perfil de correlação na proximidade das disparidades correctas é mais pronunciado.

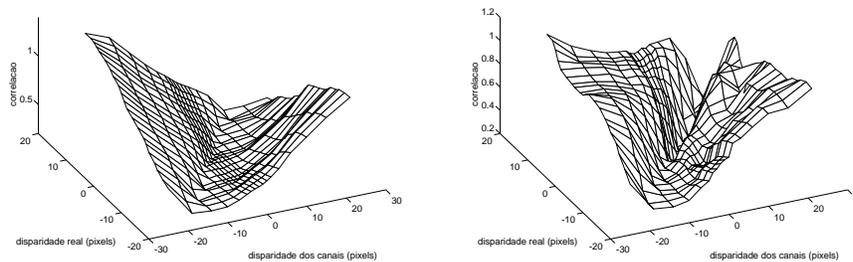


Fig. 4.8: Cálculo da correlação em diversos canais de disparidade para imagens cartesianas (esquerda) e log-polar (direita).

No caso cartesiano, esta zona apresenta uma forma quase plana, o que significa que estimativas erradas de disparidade são prováveis de acontecer. Estes resultados são coerentes com o que foi observado para a evolução da função de correlação com o ângulo de vergência.

Um aspecto interessante desta estratégia de controlo tem a ver com a analogia que é possível estabelecer, segundo determinados autores, com o funcionamento do sistema visual humano. De entre as diversas teorias existentes sobre o controlo dos movimentos de

vergência, em [58] introduz-se a *teoria do modo duplo*¹, que considera que os movimentos de vergência são compostos por duas fases: inicialmente, quando a disparidade retinal é suficientemente grande, a resposta consiste numa sequência de movimentos bruscos para diminuir rapidamente o valor da disparidade; numa segunda fase procede-se a um ajuste fino do ângulo de vergência através de uma resposta mais suave. Por outro lado, em [50] propõe-se uma teoria baseada num número discreto de canais de disparidade que são activados em função do nível de disparidade presente nas imagens.

Relativamente a esta última teoria, cada canal, quando activado, produz uma resposta com uma característica do tipo passa-baixo, com diferentes ganhos e constantes de tempo (ver Figura 4.9). Na nossa abordagem optou-se por utilizar o valor da disparidade do canal vencedor directamente como entrada do sistema de controlo, o que corresponde a dotar cada canal com características dinâmicas idênticas.

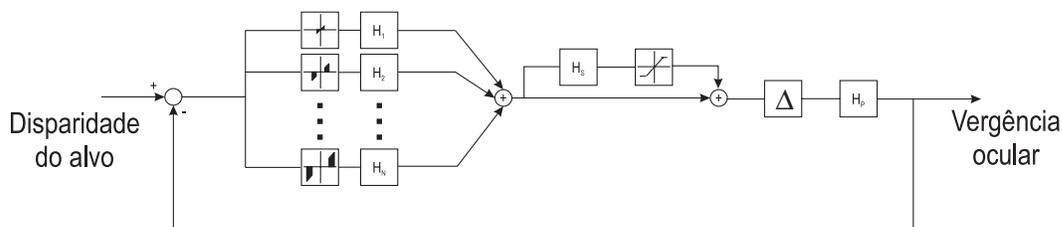


Fig. 4.9: Diagrama de blocos do modelo de controlo de vergência humano, segundo M.Pobuda e C.Erkelen [50]. Uma parte essencial do modelo é o processamento de disparidade em vários canais paralelos, cada um deles sensível a intervalos de disparidade específicos e respostas com características passa-baixo distintas. Cada bloco H_i representa a função de transferência do canal i .

4.4 Experiências e resultados

Os algoritmos de controlo de vergência propostos foram implementados na base experimental Medusa [56]. Ambos os processos correm em tempo-real, ao ritmo de amostragem do sinal de vídeo (25 Hz).

¹em inglês, *dual mode theory*.

Foram executadas algumas experiências com a finalidade de testar o desempenho dos algoritmos, sendo utilizados variados objectos localizados a diversas distâncias da cabeça robótica. Na grande maioria das situações estáticas, e para condições iniciais dentro de determinados limites, ambas as abordagens conseguem atingir os ângulos de vergência correctos.

No algoritmo de controlo por optimização global foram definidos tantos canais quanto os permitidos, de modo a não exceder o tempo de processamento disponível em cada ciclo de amostragem. Na implementação actual são utilizados 17 canais de disparidade².

Para avaliar a resposta do sistema a entradas do tipo escalão, colocou-se um objecto a uma determinada distância das câmaras que, posteriormente se moveu para duas localizações mais próximas e depois se retirou-se do campo visual. A Figura 4.10 apresenta a evolução temporal do ângulo de vergência e do sinal de controlo, e a Figura 4.11 mostra duas imagens obtidas pela câmara esquerda no decorrer da experiência.

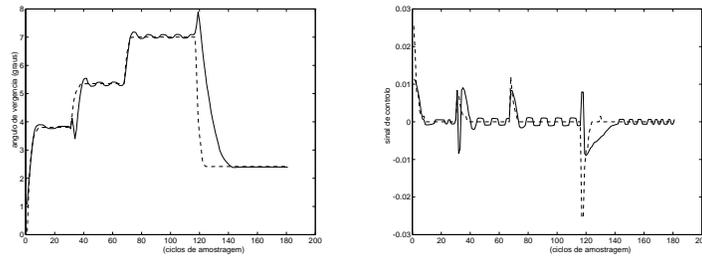


Fig. 4.10: Resposta do sistema de controlo de vergência a estímulos do tipo escalão. Linha contínua: controlo por optimização local. Linha interrompida: controlo por optimização global.

As câmaras moveram-se para vergir no objecto sempre que este mudou de posição e finalmente fixaram-se no fundo da sala. Um controlador proporcional em malha fechada foi utilizado nos dois algoritmos propostos. As unidades da escala temporal correspondem a um período de amostragem (40 ms).

Por observação da Figura 4.10, a principal distinção a efectuar entre os dois algoritmos de controlo é relativa à oscilação nas proximidades do ângulo correcto de vergência, exibida pelo controlo por optimização local. Isto é devido ao facto de este algoritmo se basear

²para o conjunto de valores $0, \pm 1, \pm 2, \pm 3, \pm 5, \pm 8, \pm 12, \pm 17, \pm 23$ pixels



Fig. 4.11: Imagens típicas obtidas durante as experiências de vergência

no sinal da diferença temporal entre valores da correlação em instantes consecutivos. Para obter este sinal, as câmaras têm que se mover permanentemente, mesmo quando é atingido o ângulo de vergência correcto. Este facto restringe o valor do ganho proporcional do controlador, limitando a rapidez do sistema, uma vez que ganhos elevados produzem oscilações de maior amplitude. Quando as câmaras vergem no fundo da cena, as oscilações não são visíveis na figura, dado terem muito menor amplitude, porque o sinal de controlo depende do valor da correlação que é muito baixo nesta situação (Figura 4.12).

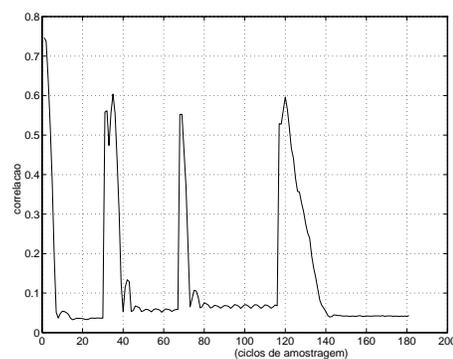


Fig. 4.12: Evolução do valor da correlação durante a experiência efectuada.

O controlo por optimização global demonstra um desempenho melhor, quer ao nível do tempo de estabelecimento quer ao nível da estabilidade na convergência. Os ângulos de

vergência correctos são atingidos após um curto transitório, e uma vez atingidos, o sistema mantém-se estável. Por outro lado, verificou-se que a amplitude das disparidades iniciais a partir das quais o sistema converge, é muito superior no caso do controlo por optimização global. Isto depende, obviamente, do alcance do conjunto de canais de disparidade — o que é fácil de definir — enquanto no caso do controlo por optimização local, a dependência advém da largura da região de convergência da função de correlação³, que pode ser da ordem dos 5 graus.

Ao nível das estratégias de controlo a utilizar, o algoritmo de optimização global revela-se mais simples. O valor estimado por este processo pode ser directamente utilizado como entrada de um controlador simples, do tipo PID. Na experiência apresentada, utilizou-se um controlador proporcional, pelo que, à parte um factor de escala, o sinal de controlo corresponde à estimativa de disparidade obtida. (Figura 4.13).

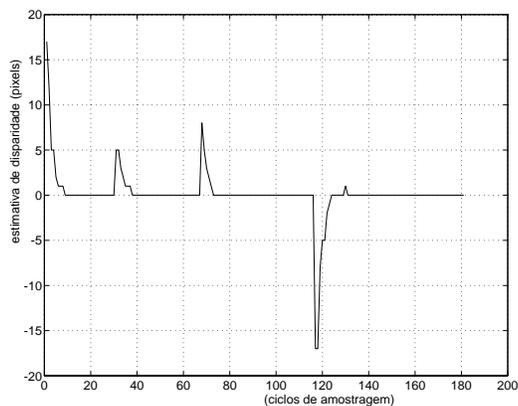


Fig. 4.13: Estimativas de disparidade obtidas ao longo da experiência efectuada.

Outra importante vantagem do controlo por optimização global consiste num melhor desempenho em situações dinâmicas. Efectuaram-se algumas experiências com um objecto que se move em profundidade a velocidade constante, conforme a Figura 4.14.

A evolução do ângulo de vergência na fixação desse objecto é apresentada na Figura 4.15, para os dois algoritmos de controlo propostos. Observa-se uma resposta melhor comportada por parte do controlo por optimização global; a outra estratégia chega a exibir instabilidade na resposta. Isto acontece porque o algoritmo de controlo por optimização

³a região de convergência é composta pela reunião das regiões monótonas adjacentes ao mínimo.

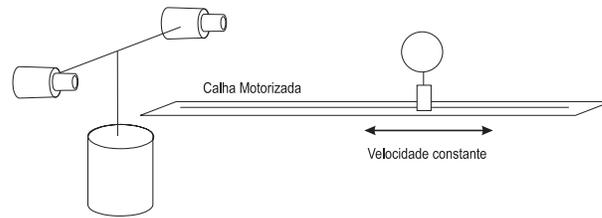


Fig. 4.14: Aparato experimental utilizado para testes de vergência com alvos móveis.

local requer que a função de correlação não varie muito de instante para instante, para que o processo de *gradient descent* funcione. Dependendo do movimento dos objectos presentes no campo visual, esta condição pode não se verificar e conduzir a situações de instabilidade. Outros resultados relativos ao mesmo sistema de vergência poderão ser consultados em [12].

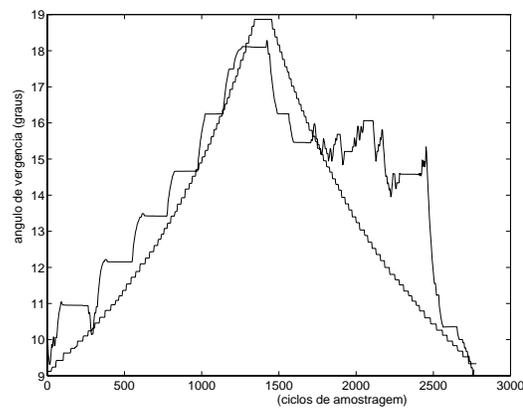


Fig. 4.15: Resposta dos algoritmos de controlo propostos a um estímulo de velocidade constante. O controlo por optimização global mostra um desempenho mais adequado a situações dinâmicas.

5. Seguimento

Em alguns sistemas de seguimento visual propostos [39, 31, 36], considera-se o problema de seguimento como um problema monocular. Mesmo em sistemas binoculares muitas vezes o controlo do movimento das câmaras é feito independentemente a partir da informação visual adquirida por cada uma [52].

Conseguindo resolver problemas de seguimento monocular activo relativos a um mesmo objecto com cada uma das câmaras de um sistema *stereo*, soluciona-se também o problema binocular. No entanto, considerando o sistema binocular como um todo e utilizando pistas binoculares para efectuar o seguimento, é possível aproveitar as relações geométricas específicas do sistema e a informação conjunta presente nas imagens *stereo* para tornar mais simples e fiável o processo de seguimento.

Tendo em conta a geometria do sistema, mostrou-se que adoptando fixação frontal é possível decompor os movimentos de seguimento em duas componentes separáveis: movimentos em profundidade e movimentos direccionais. Vimos no Capítulo 4 que através da utilização de uma pista binocular como a disparidade é possível compensar as alterações de profundidade do alvo. Portanto, para solucionar o problema do seguimento binocular resta encontrar formas de efectuar a compensação da componente direccional do movimento do alvo.

5.1 Hipótese de vergência permanente

Ao longo deste capítulo vamos considerar a **hipótese de vergência permanente**. Admite-se que o processo de vergência é perfeito, ou seja, compensa instantaneamente as variações de profundidade do alvo. Esta hipótese traz duas implicações importantes

no processo de seguimento.

Em primeiro lugar, como podemos decompor qualquer movimento do alvo em componentes de profundidade e direcção, o processo de seguimento simplifica-se e resume-se a compensar a orientação do ponto fixação relativamente ao alvo. Por outras palavras, utilizando as coordenadas esféricas para representar os pontos no espaço (ver Figura 2.5), a hipótese anterior significa que as coordenadas do alvo e do ponto de fixação segundo ρ são igualladas instantaneamente, restando corrigir as coordenadas ϕ e γ , que na cabeça robótica Medusa estão directamente relacionadas com movimentos puros de *pan* e de *tilt*.

Em segundo lugar, o alvo encontra-se permanentemente no horóptero do observador (Secção 2.4.3), ou seja, zonas semelhantes do alvo projectam-se em localizações semelhantes nas imagens (Figura 5.1).

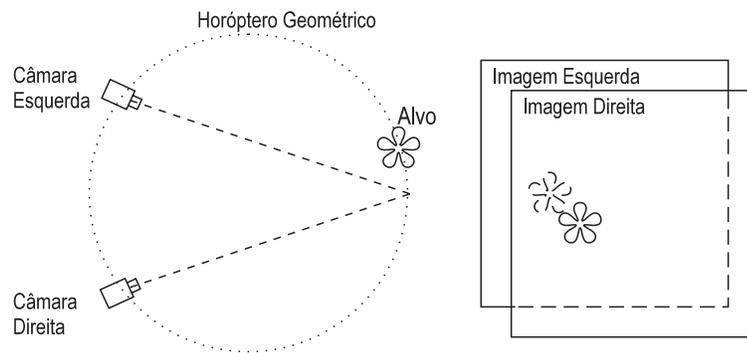


Fig. 5.1: Objectos localizados no horóptero têm disparidades nulas.

Este facto tem duas vantagens importantes:

1. facilita a discriminação entre pontos que pertencem ao objecto sob vergência e pontos que pertencem a outros objectos ou ao fundo;
2. permite integrar a informação *stereo* do alvo numa representação comum às duas imagens.

A primeira vantagem é bastante importante uma vez que um dos principais problemas em tarefas de seguimento é precisamente a segmentação do alvo relativamente ao fundo. Este assunto será aprofundado no próximo capítulo e, por enquanto, vamos admitir que

existe um processo visual que permite fazer essa distinção e que será designado por **processo de segmentação**. A segunda permite a utilização de apenas uma imagem com vista à obtenção de medidas úteis para o seguimento.

5.1.1 Representação ciclópica

Sob a hipótese de vergência permanente o alvo de interesse ocupa localizações muito semelhantes nas duas imagens. Numa situação ideal, todos os pontos do alvo teriam disparidade nula, pelo que as duas imagens conteriam exactamente a mesma informação sobre o alvo. Esta redundância pode ser aproveitada para condensar a informação obtida pelo sistema *stereo*.

Como qualquer das imagens, isoladamente, contém toda a informação obtida sobre o alvo, bastaria considerar uma delas para efectuar as medições necessárias para o controlo do seguimento do alvo. No entanto, raramente todos os pontos do objecto têm disparidade nula. Devido ao volume do objecto, às diferentes perspectivas obtidas por cada uma das câmaras e a imperfeições no processo de vergência, em geral as disparidades não são nulas.

A noção de **representação ciclópica** [45] é utilizada para descrever o modo como as duas imagens obtidas por um sistema *stereo* se combinam para dar origem a uma única representação do espaço visual. Podemos admitir que ambas as imagens contribuem de igual forma para a criação de uma representação comum, embora no sistema visual humano isto não aconteça sempre: muitas vezes um dos olhos é mais dominante, influenciando mais a percepção que temos do mundo envolvente. No contexto desta tese, consideramos um modelo muito simples de representação ciclópica, com igual contribuição das duas imagens. A **imagem ciclópica** I_c é definida como a média das imagens das câmaras esquerda I_e e direita I_d , em cada ponto:

$$I_c = \frac{I_e + I_d}{2}$$

Será a partir desta representação que iremos obter estímulos visuais para o controlo dos movimentos de seguimento. As restantes secções deste capítulo abordam os seguintes problemas:

- qual a informação útil para o controlo do seguimento?

- como utilizar essa informação para controlar os movimentos de seguimento?
- como obter essa informação a partir das sequências de imagens recolhidas pelo sistema?

5.2 Estímulos de seguimento

Admitindo a hipótese de vergência permanente e uma representação ciclópica para as imagens, o problema de seguimento traduz-se em manter a posição do alvo o mais próximo possível do centro da imagem ciclópica, ao longo do tempo.

No sistema visual humano o controlo oculomotor tenta, de uma forma contínua no tempo, minimizar a velocidade relativa entre o objecto e o ponto de fixação ocular por intermédio de movimentos de seguimento lento. Por vezes, esta estratégia de controlo não é suficiente para manter o objecto de interesse no centro das retinas. Existem, no caso geral, erros no seguimento da posição do alvo, que poderão ainda aumentar devido a imperfeições no processo de seguimento em velocidade.

Para compensar estes erros são utilizados os movimentos micro-sacádicos. Quando a posição do objecto na retina excede determinados limites é desencadeado um movimento micro-sacádico que tenta atraí-la de novo para o centro da retina. Estes movimentos utilizam como estímulo o erro de posição retinal e, ao contrário dos movimentos de seguimento lento, ocorrem em instantes discretos no tempo. Para além disso, quando acontecem inibem o movimento de seguimento lento e a medição do escorregamento retinal [53].

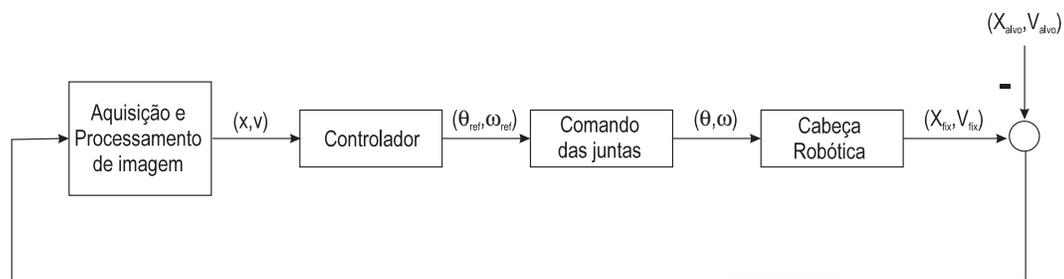
Identificam-se, assim, os dois estímulos fundamentais para o comportamento de seguimento: o escorregamento retinal e o erro de posição retinal. Em seguida analisa-se como a partir destes estímulos podemos controlar os movimentos da cabeça robótica Medusa.

5.3 Controlo de seguimento

Nesta secção aborda-se o problema de minimizar a posição e a velocidade relativa entre o ponto de fixação ocular e o alvo de interesse, de modo a efectuar o seu seguimento. Isto corresponde a controlar a posição e velocidade do alvo no plano da imagem ciclópica

para valores próximos de zero. Para atingir este objectivo, o sistema de controlo terá que comandar de forma adequada o movimento das juntas da cabeça robótica, utilizando para isso estimativas do movimento do alvo no plano de imagem, obtidas através de sensoriamento visual. Portanto, para projectar o sistema de controlo, há que conhecer a relação existente entre o movimento da cabeça robótica e o movimento do alvo no plano de imagem.

O diagrama da Figura 5.2 representa esquematicamente o sistema global a controlar, com o controlador integrado segundo uma arquitectura de controlo por retroacção.



- (X_{fix}, V_{fix}) - Posição e velocidade do ponto de fixação
- (X_{alvo}, V_{alvo}) - Posição e velocidade do alvo
- (x, v) - Posição e velocidade relativa do alvo no plano de imagem
- $(\theta_{ref}, \omega_{ref})$ - Referências de posição e velocidade angular para os motores das juntas
- (θ, ω) - Posição e velocidade angular dos motores das juntas

Fig. 5.2: Representação global do sistema.

Como se pode observar, a relação existente entre o movimento desejado para a cabeça $(\theta_{ref}, \omega_{ref})$ e as estimativas do movimento do alvo no plano de imagem (x, v) sofre a influência de diversos sub-sistemas:

- O sistema de **comando das juntas**, que recebe as referências do controlador e tenta efectuar um seguimento destas referências o mais fielmente possível. Este sistema inclui também a dinâmica electro-mecânica da cabeça robótica.
- A **cinemática do sistema robótico**, que descreve a relação entre os ângulos das juntas e o ponto de fixação visual no espaço.
- O sistema de **aquisição e processamento de imagem**, que transforma os movimentos relativos do alvo em medidas de movimento no plano de imagem, e constitui

a componente de **sensoriamento** do nosso sistema.

O sistema de comando das juntas, é composto por um sistema de controlo interno com ritmos de amostragem muito superiores aos ritmos atingíveis pelo ciclo de controlo que envolve a visão (cerca de 20x) [61]. Assim, visto à escala do sistema de controlo visual, podemos considerar que as referências introduzidas no comando das juntas se propagam quase imediatamente para a saída do motor. Podemos por isso desprezar a dinâmica electro-mecânica dos motores e da forças inerciais envolvidas na movimentação da cabeça robótica. O modelo a adoptar para o sistema a controlar considerará apenas as relações envolvidas na cinemática da cabeça robótica e no sensoriamento das grandezas de posição e velocidade do alvo.

5.3.1 Cinemática e sensoriamento

Adoptando a hipótese de vergência permanente, é possível compensar os deslocamentos do alvo simplesmente com movimentos de *pan* e de *tilt*. O modo como efectuar esses movimentos a partir do deslocamento do alvo no plano da imagem ciclópica, será discutido em seguida.

Neste trabalho não se pretende chegar a uma perfeita descrição matemática da relação entre as grandezas comandadas e as grandezas estimadas, i.e. as relações geométricas entre o movimento dos ângulos das juntas e o movimento do alvo no plano de imagem. É certo que com uma descrição matemática completa e precisa dessa relação seria possível atingir seguimentos bastante eficazes. No entanto, uma descrição deste tipo é demasiado complexa e existem sempre erros de modelização, quer na medição de parâmetros do modelo, quer pela omissão de certas propriedades do modelo, que poderiam conduzir degradações sérias no desempenho do sistema.

A abordagem tomada baseia-se na linearização dessas relações para pequenos desvios do alvo relativamente à sua posição corrente. Um sistema de controlo clássico em cadeia fechada terá como função utilizar essas relações lineares e garantir que o alvo se mantém em localizações próximas do centro da imagem ciclópica, onde os erros cometidos na linearização são desprezáveis. Isto permite simplificar bastante a análise do problema e, ainda assim, assegurar bons seguimentos do alvo.

Vamos começar por analisar que tipo de trajectória no plano de imagem é efectuada por um ponto estático no espaço para rotações puras de *tilt* e de *pan*, tentando encontrar uma decomposição lógica para estes movimentos. Para isso, começamos por adoptar uma representação comum para as imagens, como se fossem obtidas por uma câmara fictícia colocada sobre o ponto médio da distancia interocular e orientada segundo o eixo ciclópico. Adoptamos também o modelo de formação de imagem baseado na projecção perspectiva (Equação (2.4.1)). Ficamos, assim, com a geometria ciclópica representada na Figura 5.3.

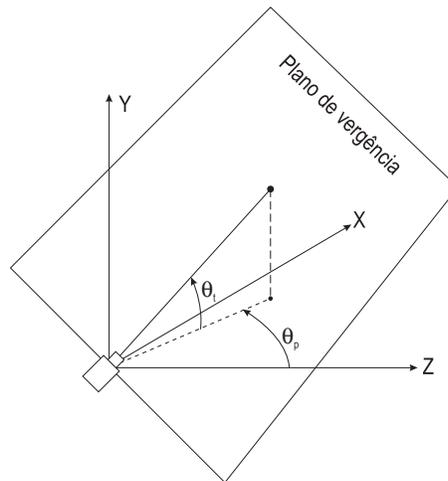


Fig. 5.3: Geometria ciclópica.

Embora a representação adoptada só constitua uma boa aproximação para ângulos de vergência pequenos, permite concluir que perante rotações das juntas de *pan* e de *tilt* as trajectórias no plano de imagem descritas por pontos estáticos no espaço visual são, de um modo geral, bastante complexas.

As rotações puras de *tilt* não provocam trajectórias puramente verticais a não ser para pontos sobre a recta vertical que passa pelo centro da imagem (Figura 5.4). Sob rotações puras de *pan*, mesmo os pontos originalmente sobre o centro da imagem só descrevem trajectórias horizontais se o ângulo de *tilt* for nulo (Figura 5.5).

Em todo o caso, por análise das figuras anteriores, pode-se concluir que, para pontos próximos da origem e ângulos de *tilt* baixos, as trajectórias devidas a movimentos de *tilt* e de *pan* podem ser localmente aproximadas por rectas verticais e horizontais, respectivamente. Nestas circunstâncias, admite-se também que variações semelhantes de ângulos

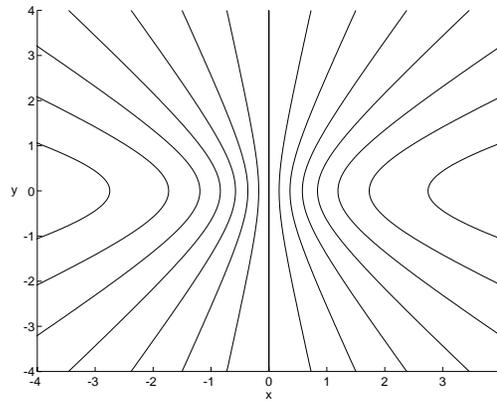


Fig. 5.4: Trajectórias de diversos pontos estáticos no espaço, induzidas por rotações puras de *tilt*.

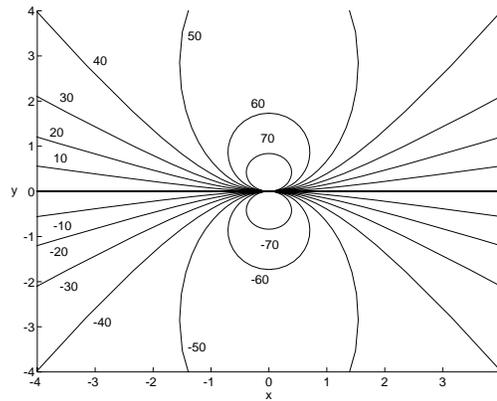


Fig. 5.5: Trajectórias de um ponto estático no espaço (originalmente no centro da imagem) induzidas por rotações puras de *pan*. São apresentadas trajectórias para diversos ângulos de *tilt* (em graus).

provocam variações semelhantes de posição no plano de imagem. Conjugando estas duas aproximações, ficamos com um modelo linear local para o sistema em análise:

$$\begin{cases} \delta\theta_{pan} \simeq \gamma_x x \\ \delta\theta_{tilt} \simeq \gamma_y y \end{cases} \quad (5.1)$$

em que γ_x e γ_y são constantes dependentes dos parâmetros intrínsecos do sistema.

Experimentalmente, estas relações lineares revelaram-se satisfatórias. Em situações estáveis de seguimento, a posição do alvo não se afasta muito do centro da imagem. Por outro lado, a junta de *tilt* não excede, por construção, cerca de 40 graus de elevação. Assim, não se cometem erros apreciáveis ao efectuar a linearização do modelo.

A relação entre a variação de velocidade angular (δw) aplicada aos motores e o vector de velocidade do alvo no plano de imagem $(v_x, v_y) = \left(\frac{dx}{dt}, \frac{dy}{dt}\right)$, pode ser obtida de uma forma análoga:

$$\begin{cases} \delta w_{pan} \simeq \gamma_u v_x \\ \delta w_{tilt} \simeq \gamma_v v_y \end{cases} \quad (5.2)$$

em que γ_u e γ_v são, de novo, constantes dependentes dos parâmetros intrínsecos do sistema.

De acordo com a análise anterior, pequenos deslocamentos do alvo poderão ser compensados através de movimentos das juntas relacionados linearmente com as medidas no plano de imagem. Para além disso, o modelo linear local obtido exprime uma decomposição importante nas relações entre os estímulos sensoriais e os movimentos das juntas. Os movimentos puros *pan* são relacionados com as componentes dos estímulos segundo a coordenada cartesiana x , enquanto os movimentos puros de *tilt* relacionam-se com as componentes segundo y . Por este motivo, os processos sensoriais a ser descritos na próxima secção tentarão obter estimativas da posição e velocidade do alvo no plano de imagem segundo as duas coordenadas cartesianas, apesar das imagens utilizadas serem representadas na geometria log-polar.

5.4 Estimação de movimento

Na estimação da posição e da velocidade do alvo no plano de imagem vamos admitir a existência de um **processo de segmentação** que classifique os pontos que pertencem ao

alvo e os pontos do restante campo visual. Em qualquer das medidas a estimar, serão apenas considerados os pontos da imagem pertencentes ao alvo. O projecto do processo de segmentação é um assunto abordado no próximo capítulo.

5.4.1 Erro de posição

Um dos estímulos a utilizar no controlo dos movimentos de seguimento é dado directamente pela posição cartesiana do alvo no plano de imagem (ver Equação (5.1)). Portanto, apesar da representação das imagens numa geometria log-polar, pretendemos obter as coordenadas cartesianas da posição do alvo.

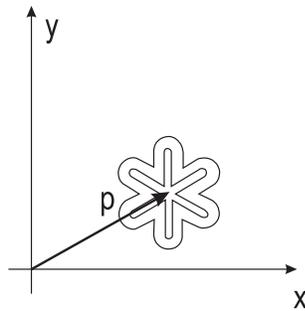


Fig. 5.6: Erro de posição retinal.

O erro de posição retinal é definido como a distância entre a posição do objecto na imagem e as coordenadas do centro da imagem (ver Figura 5.6), ou seja:

$$\mathbf{p} = (x, y)$$

Para um objecto pontual, esta definição não acarreta nenhuma ambiguidade uma vez que a posição do objecto na imagem é perfeitamente bem definida, mas para objectos não pontuais isso não acontece. Nesse caso é habitual adoptar a localização do centroide da projecção do objecto para representar a sua posição na imagem.

Admitindo a existência de um processo fiável de segmentação do objecto, então o problema da estimação da posição do objecto pode tornar-se bastante simples: por exemplo através do cálculo do centroide da zona da imagem que corresponde ao alvo. No entanto, os métodos conhecidos de segmentação apresentam geralmente elevados níveis de ruído podendo, em certas condições, perder bastante qualidade. Assim, numa tentativa

de controlar a qualidade da estimação de posição do alvo, podemos calcular também os valores da variância da zona segmentada. Estes valores fornecem informação útil sobre a fiabilidade do resultado.

Para a estimação do centroide ($\bar{\mathbf{p}}$) e da variância (Σ) da zona segmentada utilizam-se as expressões conhecidas:

$$\bar{\mathbf{p}} = \frac{\iint_{x,y} I(x,y) \mathbf{p} \, dx \, dy}{\iint_{x,y} I(x,y) \, dx \, dy} \quad \text{e}$$

$$\Sigma = \frac{\iint_{x,y} I(x,y) \mathbf{p} \mathbf{p}^T \, dx \, dy}{\iint_{x,y} I(x,y) \, dx \, dy} - \bar{\mathbf{p}} \bar{\mathbf{p}}^T$$

com:

$$\bar{\mathbf{p}} = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} \quad \text{e} \quad \Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix}$$

Relembremos, entretanto, que as imagens obtidas pelo observador artificial são representadas na geometria log-polar. Para calcular medidas cartesianas a partir de imagens log-polar há que introduzir no cálculo anterior a transformação de coordenadas inversa (Equações (A.4) e (A.6)). Assim, se $I'(\xi, \eta)$ for a imagem isomorfa de $I(x, y)$ pela transformação de coordenadas log-polar, então os valores cartesianos do centroide e da variância da posição do alvo vêm dados por:

$$\bar{\mathbf{p}} = \frac{\iint_{\xi,\eta} I'(\xi,\eta) \mathbf{T}^{-1}(\xi,\eta) |\mathbf{J}^{-1}(\xi,\eta)| \, d\xi \, d\eta}{\iint_{\xi,\eta} I'(\xi,\eta) |\mathbf{J}^{-1}(\xi,\eta)| \, d\xi \, d\eta} \quad \text{e}$$

$$\Sigma = \frac{\iint_{\xi,\eta} I'(\xi,\eta) \mathbf{T}^{-1}(\xi,\eta) \mathbf{T}^{-T}(\xi,\eta) |\mathbf{J}^{-1}(\xi,\eta)| \, d\xi \, d\eta}{\iint_{\xi,\eta} I'(\xi,\eta) |\mathbf{J}^{-1}(\xi,\eta)| \, d\xi \, d\eta} - \bar{\mathbf{p}} \bar{\mathbf{p}}^T$$

em que \mathbf{T}^{-T} é o vector que resulta da transposição de \mathbf{T}^{-1} .

Considerando imagens contínuas, os cálculos efectuados com imagens cartesianas ou log-polar são idênticos. No caso de imagens discretas isso não acontece devido aos erros de discretização introduzidos. Este facto não é grave desde que, como se espera, o alvo se mantenha próximo do centro das imagens, onde a resolução das imagens log-polar é elevada e os erros de discretização são baixos.

5.4.2 Escorregamento retinal

A forma mais directa de medir a velocidade do alvo na retina consiste em diferenciar o valor da sua posição ao longo do tempo. Esta estratégia não tem resultados muito satisfatórios quando os valores da posição estão afectados de níveis de ruído consideráveis¹

Outra forma bastante conhecida para estimação de velocidades retiniais é o cálculo do **fluxo óptico** [34]. O fluxo óptico é definido como o deslocamento aparente dos padrões de brilho na imagem e o seu cálculo obtém para cada ponto da imagem (x, y) , um vector (v_x, v_y) que representa a velocidade estimada para esse ponto.

A integração da informação do fluxo óptico para cada ponto pertencente ao objecto torna o processo de estimação do escorregamento retinal mais imune ao ruído, o que vai provocar importantes melhorias no controlo do movimento de seguimento lento.

Existem diversos métodos de cálculo do fluxo óptico: métodos diferenciais [33], métodos de procura por correlação de áreas [10] e baseados em fase [28], entre outros. Os métodos diferenciais são os de cálculo mais rápido e serão, por isso, escolhidos nesta fase do trabalho. Baseiam-se apenas na aplicação de filtragens espaciais e temporais às imagens e, ao contrário de outros métodos, não necessitam de efectuar procura ao longo das imagens, que é um processo computacionalmente exigente.

Cálculo do fluxo óptico

Dentro dos métodos diferenciais de cálculo de fluxo óptico existem várias abordagens: **(i)** o **fluxo normal** [34], que representa em cada ponto a componente da velocidade na direcção do gradiente da imagem; **(ii)** métodos de primeira ordem com termos de regularização [33], que calculam o vector de fluxo óptico completo à custa da introdução de restrições de suavidade para o fluxo e **(iii)** métodos de segunda ordem [10], que calculam o fluxo óptico completo através da introdução de segundas derivadas espaciais da imagem.

Neste trabalho efectua-se o cálculo do fluxo normal. A equação base para o cálculo do fluxo normal é a conhecida restrição:

$$I_x v_x + I_y v_y + I_t = 0 \quad (5.3)$$

¹a operação de diferenciação amplifica substancialmente o ruído.

em que $(I_x, I_y) = \nabla \mathbf{I}$ é o gradiente espacial da imagem e I_t é a derivada temporal. O vector de fluxo óptico (v_x, v_y) é definido em cada ponto da imagem e corresponde ao deslocamento desse ponto, admitindo que o seu brilho se mantém constante ao longo do tempo² [33].

O fluxo normal (v_x^\perp, v_y^\perp) é a projecção do fluxo óptico na direcção do gradiente da imagem e constitui a única componente do fluxo observável a partir de medições locais, devido ao problema da abertura [34]. O seu cálculo pode ser efectuado por:

$$\begin{cases} v_x^\perp = -\frac{I_t}{I_x^2 + I_y^2} I_x \\ v_y^\perp = -\frac{I_t}{I_x^2 + I_y^2} I_y \end{cases}$$

expressão que se obtém a partir de (5.3).

O cálculo do fluxo normal é substancialmente mais rápido e estável do que o do fluxo óptico completo, obtido pelos métodos de primeira e segunda ordem atrás referidos. Embora constituído por informação parcial do fluxo óptico, o fluxo normal é com frequência suficiente para a resolução de várias tarefas visuais [54]. Pelas suas características de rapidez, simplicidade, e robustez, adequa-se perfeitamente ao requisitos de funcionamento em tempo real e de fiabilidade pretendidos para o nosso sistema.

Fluxo óptico em imagens log-polar

Mais uma vez há que ter em conta que as imagens utilizadas neste trabalho são representadas na geometria log-polar. É necessário relacionar as medidas de fluxo obtidas sobre as imagens log-polar com as respectivas medidas cartesianas que serão utilizadas no controlo dos movimentos de seguimento.

Seja $\mathbf{v}_l = (v_\xi, v_\eta)$ o vector de fluxo óptico obtido a partir das imagens log-polar e $\mathbf{v}_c = (v_x, v_y)$ o vector de fluxo óptico nas correspondentes imagens cartesianas. Dado que:

$$\begin{cases} v_x = \frac{dx}{dt} \\ v_y = \frac{dy}{dt} \end{cases} \quad \text{e} \quad \begin{cases} v_\xi = \frac{d\xi}{dt} \\ v_\eta = \frac{d\eta}{dt} \end{cases}$$

então, por simples cálculo diferencial, obtêm-se as relação entre estes dois vectores:

$$\mathbf{v}_c = \mathbf{J}^{-1} \mathbf{v}_l \tag{5.4}$$

²hipótese de brilho constante ou, em inglês, *brightness constancy assumption*.

em que \mathbf{J}^{-1} é a matriz jacobiana inversa da transformação de coordenadas log-polar, dada pela Equação (A.5).

Estimação da velocidade do alvo

O processo de estimação da velocidade retinal do alvo baseia-se na integração das medidas de fluxo normal em pontos pertencentes ao objecto. O efeito de medições erradas ou ruidosas dilui-se quando são incluídas muitas medições no processo de estimação. Este torna-se, assim, mais tolerante a falhas e produz estimativas mais estáveis do que a simples diferenciação da posição do objecto ao longo do tempo.

Dois métodos alternativos foram testados. O primeiro, mais simples, calcula apenas a média dos vectores de fluxo normal e não supõe qualquer modelo de movimento para o objecto. O segundo, consiste num método de regressão linear e admite que velocidade no plano de imagem é constante em todos os pontos pertencentes ao objecto.

Média do Fluxo O método mais simples de integrar um conjunto de medições consiste em efectuar a média desses valores. O cálculo da média dos vectores de fluxo óptico pertencentes ao objecto, $\overline{\mathbf{v}_c} = (\overline{v_x}, \overline{v_y})$, recorre à seguinte expressão:

$$\overline{\mathbf{v}_c} = \frac{\iint_{x,y} \mathbf{v}_c \, dx \, dy}{\iint_{x,y} dx \, dy}$$

onde $\mathbf{v}_c = (v_x, v_y)$ é o vector de fluxo óptico cartesiano em cada ponto da imagem.

Mais uma vez é necessário exprimir estes valores em função das medidas efectuadas nas imagens log-polar obtidas pelo sistema. Introduzindo a Equação (5.4) e a transformação inversa (Equações (A.4) e (A.6)) as estimativas da velocidade média do alvo no plano de imagem vêm dadas por:

$$\overline{\mathbf{v}_c} = \frac{\iint_{\xi,\eta} \mathbf{J}^{-1}(\xi, \eta) \, \mathbf{v}_1 \, |\mathbf{J}^{-1}(\xi, \eta)| \, d\xi \, d\eta}{\iint_{\xi,\eta} |\mathbf{J}^{-1}(\xi, \eta)| \, d\xi \, d\eta}$$

Até aqui, admitimos a acessibilidade ao fluxo óptico completo mas, como vimos atrás, apenas conseguimos calcular os vectores de fluxo normal. Vamos, portanto, cometer erros, quer na direcção quer na amplitude do vector de velocidade estimado. No entanto,

verifica-se experimentalmente que, para objectos com padrões de textura sem direcções privilegiadas, a direcção do vector estimado é bastante fiável. Surgem problemas no seguimento de objectos que apresentem texturas com direcções dominantes diferentes da direcção do deslocamento, por exemplo, um objecto com listas diagonais movendo-se horizontalmente. Os erros cometidos em amplitude não são preocupantes dado que o controlo por retroacção tende a compensá-los.

Modelo de Fluxo Constante Relembremos que a equação base para o cálculo do fluxo óptico (v_x, v_y) , numa sequência de imagens $I(x, y, t)$, utilizando a técnica diferencial, é dada por:

$$I_x v_x + I_y v_y = -I_t$$

Como já foi discutido, esta restrição apenas permite calcular a componente do fluxo óptico na direcção do gradiente da imagem em cada ponto, ou seja, o fluxo normal. A indeterminação advém de, com apenas uma equação, se pretender calcular o valor das duas incógnitas v_x e v_y . Introduzindo um modelo para o fluxo óptico, torna-se possível incluir restrições que relacionem o fluxo em diversos pontos da imagem. Ficamos, assim, com um sistema determinado ou sobredeterminado, o que nos permite calcular as duas componentes do fluxo óptico ou os parâmetros do modelo adoptado.

Considerando um modelo de fluxo constante onde todos os pontos pertencentes ao objecto se movem com a mesma velocidade no plano de imagem, resulta:

$$\begin{cases} v_x(x, y) = v_x^0 \\ v_y(x, y) = v_y^0 \end{cases}$$

portanto, em cada ponto (x_i, y_i) do objecto deverá ser satisfeita a equação:

$$I_x(x_i, y_i) v_x^0 + I_y(x_i, y_i) v_y^0 = -I_t(x_i, y_i)$$

Sendo n o número de pontos da imagem pertencem ao objecto, constroi-se o sistema de equações:

$$\begin{cases} I_x(x_1, y_1) v_x^0 + I_y(x_1, y_1) v_y^0 = -I_t(x_1, y_1) \\ \dots \\ I_x(x_n, y_n) v_x^0 + I_y(x_n, y_n) v_y^0 = -I_t(x_n, y_n) \end{cases}$$

Desde que n seja superior a 2, então podemos aplicar o método dos mínimos quadrados e obter:

$$\begin{cases} v_x^0 = \frac{\sum I_y^2 \sum I_x I_t + \sum I_x I_y \sum I_y I_t}{\sum I_x^2 \sum I_y^2 - (\sum I_x I_y)^2} \\ v_y^0 = \frac{\sum I_x I_y \sum I_x I_t - \sum I_x^2 \sum I_y I_t}{\sum I_x^2 \sum I_y^2 - (\sum I_x I_y)^2} \end{cases}$$

As derivadas espaciais que surgem na expressão anterior, I_x e I_y , são as derivadas direccionais segundo as coordenadas da imagem cartesiana $I(x, y)$. Estas derivadas terão que ser calculadas através da informação contida na imagem log-polar, $I'(\xi, \eta)$, obtida pelo sistema. Por diferenciação desta imagem em cada uma das suas coordenadas podemos obter as derivadas direccionais I'_ξ e I'_η . A relação entre as derivadas espaciais nos domínios cartesiano e log-polar é dada por:

$$\begin{bmatrix} I_x \\ I_y \end{bmatrix} = \mathbf{J}' \begin{bmatrix} I'_\xi \\ I'_\eta \end{bmatrix} \quad (5.5)$$

em que \mathbf{J}' é a transposta da matriz jacobiana da transformação log-polar (Equação (A.2)).

5.4.3 Projecto do controlador de seguimento

Depois de se ter encontrado um conjunto de relações lineares locais que permite compensar pequenos deslocamentos laterais do alvo em função da sua posição e velocidade no plano de imagem, e a forma como estimar estas grandezas num sistema artificial, resta projectar um sistema de controlo que integre estes dois tipos de conhecimento para efectuar o seguimento direccionais de alvos móveis. Isto será efectuado através de controlo por retroacção com um controlador dinâmico do tipo P.I.D.

Os controladores P.I.D. são uma classe de controladores dinâmicos relativamente bem conhecida [30]. Esta classe de controladores inclui os três tipos básicos de controlo linear — controlo proporcional (P), controlo integral (I) e controlo derivativo (D) (ver Apêndice B). Numa grande parte dos problemas de controlo linear, os controladores P.I.D. conseguem cumprir satisfatoriamente o desempenho pretendido. Para além disso, as características dinâmicas do sistema em cadeia fechada podem ser modeladas por ajuste dos parâmetros proporcional, integral e derivativo, em função do modelo dinâmico do sistema a controlar.

Na Secção 5.3.1. obteve-se um modelo em regime estático para o sistema a controlar (Equações (5.1) e (5.2)). No entanto, em regime dinâmico há que ter em conta os atrasos computacionais provocados pela estimação da posição e da velocidade do alvo no plano de imagem. Assim, para o cálculo da posição do alvo necessitamos das imagens obtidas no instante anterior, o que introduz um atraso de um instante de amostragem no ciclo visual. Para a estimação da velocidade são necessárias imagens em dois instantes de amostragem anteriores³, introduzindo um atraso de dois períodos de amostragem. Tendo em conta este facto e as Equações (5.1) e (5.2) podemos obter um modelo dinâmico aproximado para o sistema a controlar:

$$\begin{cases} x(t) = \frac{1}{\gamma_x} \delta \theta_{pan}(t - T) \\ y(t) = \frac{1}{\gamma_y} \delta \theta_{tilt}(t - T) \end{cases} \quad \text{e} \quad \begin{cases} v_x(t) = \frac{1}{\gamma_x} \delta w_{pan}(t - 2T) \\ v_y(t) = \frac{1}{\gamma_y} \delta w_{tilt}(t - 2T) \end{cases} \quad (5.6)$$

onde T é um período de amostragem do ciclo visual.

O controlador deverá garantir que a posição e a velocidade do alvo no plano de imagem sejam mantidas em valores aproximadamente nulos, actuando as juntas da cabeça robótica em função dos estímulos de posição e de velocidade simultaneamente. Como foi referido na Secção 2.2.1, no sistema ocular humano esses estímulos são utilizados por dois mecanismos neuronais separados e de formas bastante distintas. Partindo do modelo dinâmico das Equações (5.6) dimensiona-se para cada junta (*pan* e *tilt*) controladores separados de posição e de velocidade adequados para garantir boas características dinâmicas de seguimento de posição e de velocidade individualmente. As saídas são ambas referidas à mesma variável de controlo (a velocidade angular da junta), combinadas de uma forma aditiva com pesos relativos adequados, permitindo que ambos os processos estejam activos simultaneamente. O dimensionamento destes controladores é feito no Apêndice B.

A aditividade das saídas de cada um dos comportamentos não é motivada biologicamente. Justifica-se pela simplicidade prática com que se soluciona o problema da integração dos dois comportamentos; só é possível neste caso porque cada um dos controladores tem saídas referidas à mesma variável de controlo (a velocidade das juntas). Assim, o sinal de controlo injectado no motor de cada junta tem em conta simultaneamente o

³o algoritmo implementado para cálculo do fluxo normal, utiliza filtragens recursivas para dois instantes de tempo.

erro de posição e o escorregamento do alvo na retina:

$$W_{ref} = a w_{pos} + (1 - a) w_{vel}$$

em que o parâmetro a é ajustado em função da confiança (fiabilidade) associada às estimativas de posição e de velocidade obtidas.

O esquema de controlo projectado encontra-se representado na Figura 5.7.

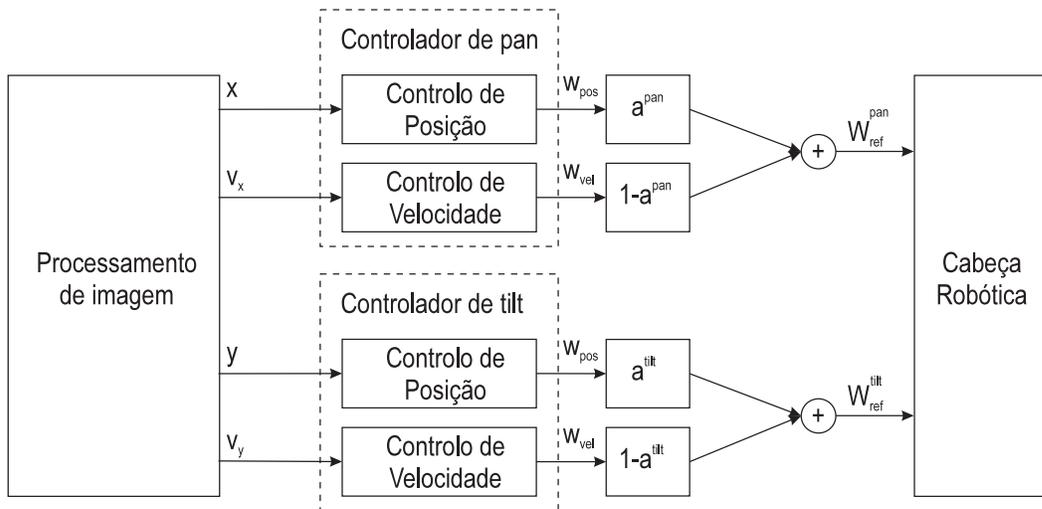


Fig. 5.7: Controlador de seguimento.

6. Integração e coordenação

Neste capítulo preocupamo-nos com a integração e coordenação dos processos de vergência e de seguimento. Em primeiro lugar, há a necessidade de dotar o processo de seguimento com um mecanismo de segmentação do alvo, uma vez que se partiu desse pressuposto no capítulo anterior. Como vamos ver, a fusão binocular proporcionada pelo processo de vergência possibilita a utilização de um método de segmentação bastante rápido e genérico. Em segundo lugar há que coordenar a geração dos movimentos oculares produzidos pela vergência e pelo seguimento.

6.1 Segmentação do alvo

Das abordagens mais utilizadas na resolução do problema de segmentação do alvo podemos distinguir entre as que são baseadas em correspondência dos níveis de brilho e as que são baseadas na segmentação do fluxo óptico. As primeiras, procuram ao longo das imagens por zonas que possam corresponder a uma descrição prévia dos objectos. Podem ser utilizadas quando o problema de seguimento se limita a um conjunto de objectos conhecidos *a priori*, ou quando o objecto a fixar é seleccionado de entre os objectos presentes na cena, por um qualquer processo de inicialização. Estes métodos são suficientemente bons para casos em que a forma do objecto não se altera muito de um instante para o outro, mas geralmente têm dificuldades em lidar com rotações e escalamentos dos objectos, entre outras deformações não rígidas.

As técnicas baseadas em segmentação de fluxo óptico conseguem lidar com alterações da forma do objecto mas, habitualmente impõem restrições quanto ao movimento relativo entre alvo e observador. Muitas vezes, para simplificar as abordagens tomadas em

problemas de seguimento, admite-se que os movimentos das câmaras são conhecidos ou que o movimento relativo do objecto a fixar segue algum modelo de possível estimação (fluxo constante, afim, etc.). Por exemplo, em problemas de seguimento passivo (câmara fixa) a detecção do alvo pode ser feita por simples diferenciação das imagens uma vez que a imagem do ambiente é estática. Com câmaras com movimento conhecido, é nalguns casos possível obter a localização dos alvos móveis por análise do fluxo óptico incoerente com o modelo adoptado [44].

No caso em que o movimento das câmaras é desconhecido ou muito complexo, é extremamente difícil detectar a localização do alvo por análise do fluxo óptico da cena. Técnicas de *clustering* de fluxo óptico [64] ou análise de resíduos na estimação do movimento próprio¹ [5, 59], podem ser utilizadas neste último caso mas os ambientes onde obtêm bons resultados são algo limitativos e não se revelam adequadas a aplicações de tempo real, pois necessitam de sequências longas de imagens. Sem admitir quaisquer restrições quanto ao movimento das câmaras ou dos alvos a fixar, é difícil estabelecer um critério para detectar a posição do alvo, a partir do fluxo óptico presente nas imagens.

Como em funcionamento normal se espera que o nosso sistema efectue um seguimento estável do alvo, poder-se-ia estabelecer um critério de selecção baseado no facto de o fluxo óptico do alvo ter valores baixos e ser elevado nos restantes pontos da imagem. No entanto, em situações transitórias pode suceder o inverso: quando um objecto, inicialmente parado se começa a mover, o seu fluxo óptico é elevado e decresce posteriormente à medida que o seguimento estabiliza, enquanto o fluxo óptico devido ao fundo tem uma evolução oposta.

Resumindo, quer as técnicas baseadas em correspondência dos níveis de brilho, que limitam as variações de forma admissíveis para o alvo, quer as técnicas baseadas em segmentação de fluxo óptico que limitam o movimento relativo do alvo, não satisfazem a generalidade pretendida para o nosso sistema. Como veremos na próxima secção, a abordagem tomada para a segmentação do alvo não utiliza qualquer técnica que imponha restrições à forma ou movimento relativo do objecto a fixar.

¹do inglês *egomotion*.

6.1.1 Fusão binocular

Quando um objecto é colocado numa posição suficientemente isolada ao longo da direcção de observação, provoca movimentos de vergência que o tendem a colocar no centro das duas retinas. Nesta situação, as zonas pertencente ao objecto apresentam disparidades aproximadamente nulas. Dá-se a **fusão binocular** [45].

O facto de os pontos pertencentes ao objecto terem disparidades baixas, ao contrário das outras zonas do espaço fora do horóptero, pode ser aproveitado para simplificar grandemente o processo de segmentação do objecto relativamente ao fundo. O algoritmo de segmentação consistirá, basicamente, em extraír do par *stereo* de imagens, as zonas correspondentes que tenham a mesma localização nas duas imagens. No entanto, existem zonas do espaço que não pertencem ao objecto e apresentam ainda disparidade nula: qualquer objecto estranho que esteja localizado no horóptero será também segmentado por este processo. De um modo geral, por este método só é possível isolar a localização do objecto de interesse se este for o único objecto do campo visual a estar presente no horóptero.

Relembramos que, para a geometria binocular do sistema robótico utilizado, pertencem ao horóptero todos os pontos localizados numa circunferência que passa pelo ponto de fixação e pelos centros ópticos das câmaras. Numa situação prática, interessa generalizar a noção de horóptero de modo a permitir que variações pequenas de profundidade do alvo sejam toleradas pelo processo de segmentação. Define-se assim o **horóptero generalizado**, como todas as localizações do espaço que produzam disparidades de valor absoluto inferior a um determinado limiar. Para uma representação cartesiana de imagens, o horóptero generalizado é limitado por duas elipses que passam pelos centros ópticos das câmaras, como se pode ver na Figura 6.1. Os objectos que se encontrem dentro deste espaço serão, em condições ideais, detectados pelo processo de segmentação.

6.1.2 Segmentação por disparidade nula

O processo de segmentação por disparidade nula consiste em detectar pontos que ocupem a mesma localização, ou localizações próximas, nas duas imagens *stereo*. É um tipo de filtragem não linear que produz uma imagem binária indicando os pontos que evidenciam

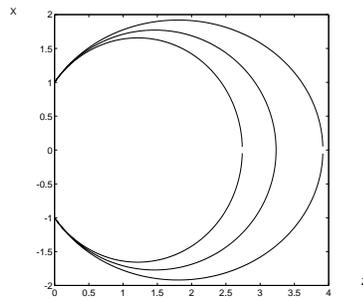


Fig. 6.1: O horóptero generalizado no plano de vergência para imagens cartesianas.

disparidades baixas.

Alguns algoritmos de segmentação por disparidade nula encontram-se documentados em [23]. Baseiam-se fundamentalmente em *matching* de características que ocupam localizações semelhantes nas duas imagens. Entre as técnicas utilizadas encontram-se: a correlação de áreas, a correspondência entre contornos verticais e a correlação de orientação e magnitude de contornos. Dado que nas imagens só existe informação relevante em zonas com textura (variações dos níveis de brilho) os conjuntos de pontos segmentados serão mais ou menos esparsos consoante a textura da cena. Por outro lado, como resultam de técnicas locais, podem dar origem a correspondências incorrectas, por exemplo quando pontos diferentes do espaço projectam características semelhantes com iguais localizações nas duas imagens.

Métodos de correlação de fase também são potencialmente utilizáveis na solução deste problema [67, 29]. Alguns têm a vantagem de analisar as imagens a várias escalas, produzindo mais correspondências correctas entre as imagens e segmentações mais densas do que os algoritmos baseados em *matching* de características, o que pode trazer vantagens para os passos seguintes de processamento. Contudo, são métodos pesados computacionalmente, o que dificulta a sua utilização em muitas aplicações.

Filtragem de disparidade nula por *matching* de contornos verticais

Neste trabalho utiliza-se um algoritmo de segmentação por disparidade nula que opta por uma abordagem de *matching* de características. É baseado num teste de semelhança aos contornos verticais em localizações idênticas nas duas imagens de um par *stereo*. A

utilização de contornos verticais é justificada pela sua fácil detecção e por conterem informação relevante sobre a disparidade horizontal. Os contornos horizontais não fornecem informação útil sobre a disparidade horizontal, uma vez que produzem ambiguidade na localização da correspondência.

A forma mais geral de implementar um detector de contornos consiste em calcular o gradiente da imagem e aplicar um limiar adequado. No caso da detecção de contornos verticais, é suficiente calcular o gradiente horizontal da imagem. A aplicação do limiar não é usada no nosso caso, dado ser desejável preservar a forma do contorno para produzir uma melhor correspondência entre as características. O teste de semelhança é feito directamente com o gradiente horizontal das imagens.

Assim, o primeiro passo do algoritmo consiste em calcular as derivadas horizontais de cada uma das imagens do par *stereo*. No segundo passo, comparam-se essas derivadas em iguais localizações nas duas imagens. As zonas que sejam semelhantes e que, simultaneamente, contenham informação fiável, são consideradas como tendo disparidade nula. Este passo consiste, portanto, em dois testes:

- o **teste de semelhança** entre as zonas da imagem;
- o **teste de qualidade** verifica se a informação contida na zona a analisar é suficiente para produzir um resultado fiável. Isto está relacionado com o valor do gradiente horizontal em cada ponto da imagem: um baixo valor absoluto revela a inexistência de contorno vertical ou um contorno demasiado fraco para poder ser utilizado no teste de semelhança.

Qualquer destes testes pode ser feito ponto a ponto ou numa área vizinha do ponto a analisar. Para o caso ponto a ponto, sendo $I^1(x, y)$ e $I^2(x, y)$ as imagens do par *stereo* original e S e Q limiares adequados, os testes podem ser feitos da seguinte forma:

- teste de semelhança — os contornos são semelhantes no ponto (x, y) se:

$$\left| \frac{\partial I^1(x, y)}{\partial x} - \frac{\partial I^2(x, y)}{\partial x} \right| < S$$

- teste de qualidade — os contornos são fiáveis se:

$$\left| \frac{\partial I^1(x, y)}{\partial x} \right| > Q \quad e \quad \left| \frac{\partial I^2(x, y)}{\partial x} \right| > Q$$

Para operações feitas numa área vizinha (A) do ponto a analisar, o teste de semelhança poderá ser feito através de uma medida de distância, por exemplo:

$$\iint_{x,y \in A} \left[\frac{\partial I^1(x,y)}{\partial x} - \frac{\partial I^2(x,y)}{\partial x} \right]^2 dx dy$$

e o teste de qualidade poderá ser feito através de uma norma do tipo:

$$\iint_{x,y \in A} \left[\frac{\partial I(x,y)}{\partial x} \right]^2 dx dy$$

Na próxima secção vamos analisar o comportamento deste algoritmo quando aplicado a imagens log-polar, estabelecendo as diferenças existentes para a geometria cartesiana.

6.1.3 Influência da representação log-polar

O algoritmo de filtragem de disparidade nula visto anteriormente pode ser aplicado, de uma forma muito semelhante, a imagens log-polar. Há, no entanto que ter em conta dois aspectos que diferem significativamente do caso cartesiano:

- a representação dos contornos verticais;
- a forma do horóptero generalizado.

Representação dos contornos verticais

O algoritmo de filtragem de disparidade nula requer informação sobre os contornos verticais presentes nas imagens cartesianas. Isto é obtido à custa do cálculo do gradiente segundo a coordenada x , que em coordenadas cartesianas representa direcções horizontais. No caso log-polar, a direcção horizontal corresponde à coordenada ξ e a direcção vertical corresponde a η . Como consequência, os contornos verticais cartesianos não podem ser obtidos simplesmente pelo cálculo do gradiente horizontal em imagens log-polar.

A relação entre os gradientes de imagem nas duas geometrias foi estabelecida anteriormente (Equação (5.5)). Particularizando para a derivada horizontal pretendida, temos:

$$I_x = \frac{\cos(\eta/\alpha)}{\rho_{\min} k^\xi \log k} I'_\xi - \frac{\sin(\eta/\alpha)}{\rho_{\min} k^\xi \alpha} I'_\eta \quad (6.1)$$

Desta forma, o gradiente necessário aos testes efectuados pelo algoritmo de segmentação pode ser obtido, para cada ponto (ξ, η) , através do cálculo das derivadas horizontais e verticais da imagem log-polar, I'_ξ e I'_η , seguido do cálculo da Expressão (6.1).

Horóptero generalizado

O horóptero generalizado foi definido, para um determinado ângulo de vergência, como o conjunto de pontos no espaço visual que apresentam uma disparidade retinal inferior a um determinado limiar. Numa geometria cartesiana, a “espessura” do horóptero diminui do centro para a periferia da imagem (ver Figura 6.1). Esta é uma forma natural de “rejeitar” (não segmentar) objectos que se encontrem no horóptero do observador mas afastados da direcção de observação.

Na geometria log-polar, as tolerâncias de profundidade que decorrem de disparidades idênticas têm uma forma substancialmente diferente, como ilustra a Figura 6.2.

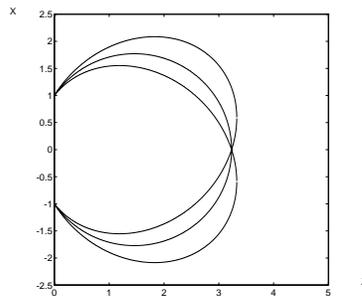


Fig. 6.2: O horóptero generalizado no plano de vergência, para imagens log-polar.

É de notar que a fronteira da zona de tolerância é constituída por elipses que passam pelos centros ópticos do sistema binocular e pelo ponto de fixação mas, ao contrário do caso cartesiano, as elipses não se encontram orientadas segundo a direcção de observação. Isto torna mais provável a segmentação de objectos indesejados que surjam em regiões periféricas das imagens.

Para evitar esta deformação do horóptero e tentar equalizar as tolerâncias de profundidade com o caso cartesiano seria necessário considerar tolerâncias de disparidade variável de ponto para ponto da imagem. No entanto, isto implicaria uma análise de imagem a várias escalas, complicando significativamente o processamento a efectuar pelo algoritmo

de segmentação. Opta-se por uma solução empírica que apresenta bons resultados práticos com reduzido esforço computacional e que consiste num processo de aumento de exigência no teste de semelhança de características, à medida que se caminha para a periferia. Isto é efectuado através de um limiar de semelhança $S(\xi)$ dependente da coordenada radial da transformação log-polar, definido como:

$$S(\xi) = \frac{s_0^{-\frac{\xi}{\xi_{\max}}+2} - s_0}{s_0 - 1} \quad (6.2)$$

em que s_0 representa o limiar inicial, $s_0 = S(0)$, e no limite periférico da imagem a função S toma valores nulos, ou seja, $S(\xi_{\max}) = 0$. Na Figura 6.3 apresenta-se a evolução da função S para alguns valores do limiar inicial.

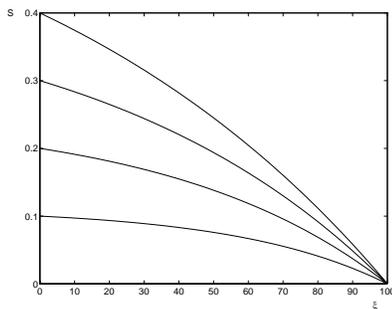


Fig. 6.3: Limiar do teste de semelhança, para alguns valores do limiar inicial.

A aplicação do limiar $S(\xi)$ faz com que os pontos da periferia só sejam segmentados se a sua semelhança for muito elevada. Isto evita segmentações de objectos que se encontrem no horóptero generalizado mas afastados da distância de fixação.

A forma exponencial inversa para a evolução deste limiar tem como motivação a evolução exponencial da disparidade cartesiana ao longo da coordenada radial, para uma disparidade log-polar constante. É fácil mostrar, que ao longo da linha horizontal que passa pelo centro da imagem cartesiana ($y = 0$), a relação entre pequenas disparidades correspondentes nas geometrias cartesiana e log-polar² é dada por:

$$d_x = e^\xi (e^{d\xi} - 1)$$

²com parâmetros $\rho_{\min} = 1$ e $k = e$.

onde d_x é a disparidade no plano cartesiano entre dois pontos cujos correspondentes log-polar têm disparidade constante d_ξ (Figura 6.4).

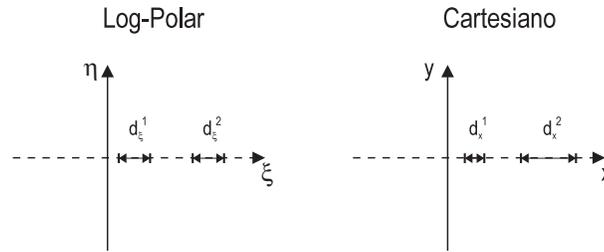


Fig. 6.4: Disparidades log-polar idênticas (d_ξ) em diferentes pontos da imagem correspondem a disparidades cartesianas (d_x) distintas.

A análise da expressão anterior revela que a disparidade cartesiana evolui de forma exponencial crescente ao longo da coordenada ξ . Como pretendemos de alguma forma inverter esta tendência, atribui-se uma evolução inversa ao valor do limiar S .

6.1.4 Algoritmo de segmentação

Partindo do algoritmo básico de segmentação de disparidade nula por *matching* de contornos verticais e tendo em conta as alterações devidas ao facto de se utilizar uma geometria log-polar para representar as imagens, desenvolveu-se um novo algoritmo. Este algoritmo é aplicado a um par *stereo* de imagens log-polar, $I^1(\xi, \eta)$ e $I^2(\xi, \eta)$, e consiste nos seguintes passos:

1. Cálculo das derivadas horizontais e verticais de cada uma das imagens, obtendo-se $I_\xi^1(\xi, \eta)$, $I_\xi^2(\xi, \eta)$, $I_\eta^1(\xi, \eta)$ e $I_\eta^2(\xi, \eta)$.
2. Cálculo das derivadas segundo a coordenada cartesiana x , para cada imagem, segundo a Expressão (6.1), obtendo-se as imagens $I_x^1(\xi, \eta)$ e $I_x^2(\xi, \eta)$.
3. Aplicação do teste de qualidade a cada uma das imagens de gradiente segundo x :

$$I_q^i(\xi, \eta) = \begin{cases} 1 & \text{se } |I_x^i(\xi, \eta)| > Q \\ 0 & \text{c.c.} \end{cases}$$

Obtêm-se as imagens binárias $I_q^1(\xi, \eta)$ e $I_q^2(\xi, \eta)$.

4. Cálculo da imagem de qualidade das características:

$$I_q(\xi, \eta) = \begin{cases} 1 & \text{se } I_q^1(\xi, \eta) = I_q^2(\xi, \eta) = 1 \\ 0 & \text{c.c.} \end{cases}$$

5. Aplicação do teste de semelhança entre as imagens de gradiente segundo x :

$$I_s(\xi, \eta) = \begin{cases} 1 & \text{se } \sum_{(\xi, \eta) \in A} [I_x^1(\xi, \eta) - I_x^2(\xi, \eta)]^2 < S(\xi) \\ 0 & \text{c.c.} \end{cases}$$

O conjunto A representa uma vizinhança de dimensão fixa, centrada no ponto (ξ, η) . A função $S(\xi)$ é dada por (6.2).

6. Resultado da filtragem de disparidade nula:

$$I_0(\xi, \eta) = \begin{cases} 1 & \text{se } I_s(\xi, \eta) = I_q(\xi, \eta) = 1 \\ 0 & \text{c.c.} \end{cases}$$

Os pontos (ξ, η) com valor 1 são considerados pontos de disparidade nula.

Foram efectuadas algumas experiências, em pares de imagens log-polar, com o algoritmo atrás referido. A Figura 6.5 apresenta o resultado obtido em duas dessas situações.

Podemos observar que as zonas contendo textura suficiente e que pertencem ao objecto vergido são correctamente segmentadas. De um modo geral, em situações de vergência correcta, ou seja, quando o alvo se encontra no interior do horóptero, os resultados da segmentação são bastante satisfatórios. Quando o ponto de fixação se afasta em profundidade do alvo, surgem pontos incorrectamente classificados espalhados pela imagem, o que pode provocar dificuldades nas fases seguintes de processamento, por exemplo, na estimação da posição do alvo no plano da imagem. Por isso, é importante que o processo de vergência seja bastante rápido e preciso na convergência para o alvo.

6.2 Vergência e seguimento

Ao longo desta tese houve a preocupação de distinguir os diferentes movimentos oculares presentes no sistema visual humano e os estímulos que os controlam. O esforço de

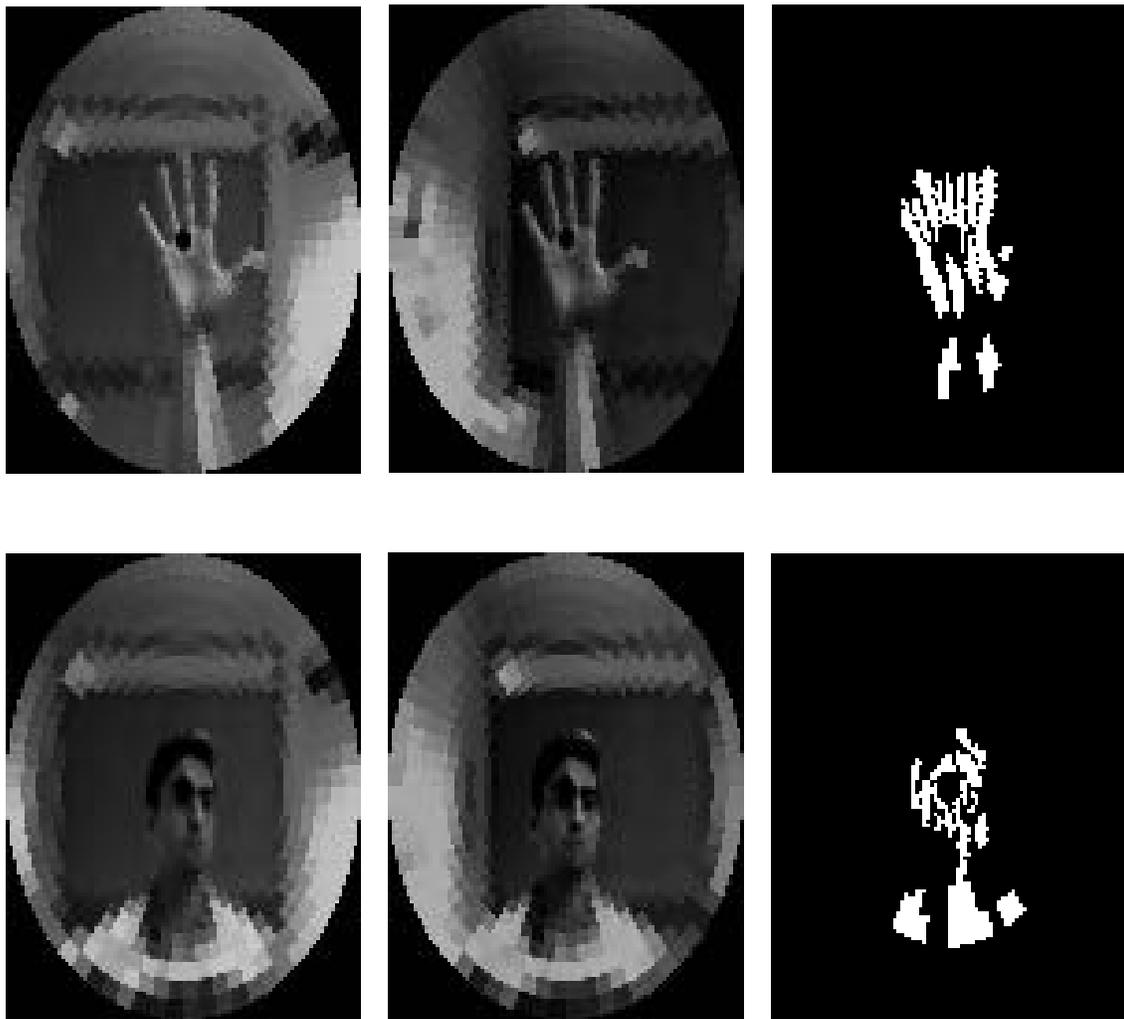


Fig. 6.5: Resultados da segmentação do alvo relativamente ao fundo, utilizando filtragem de disparidade nula sobre imagens log-polar. Da esquerda para a direita apresentam-se as imagens obtidas pelas câmaras e o resultado da segmentação, em duas situações.

adaptação destes mecanismos biológicos ao projecto de um observador artificial motivou a definição de dois processos oculomotores principais para a implementação do comportamento de fixação da atenção: a vergência e o seguimento lateral do alvo. Tendo em consideração os graus de liberdade e a estrutura mecânica da cabeça robótica Medusa, assim como condições que simplificam a análise geométrica do sistema, foi possível desacoplar os movimentos gerados por esses dois processos:

- o processo de vergência gera apenas movimentos simétricos de rotação das câmaras;
- o processo de seguimento lateral gera apenas movimentos das juntas de *pan* e de *tilt*.

Poderíamos, assim, considerar a coordenação como uma simples aditividade dos processos, como representado na Figura 6.6.

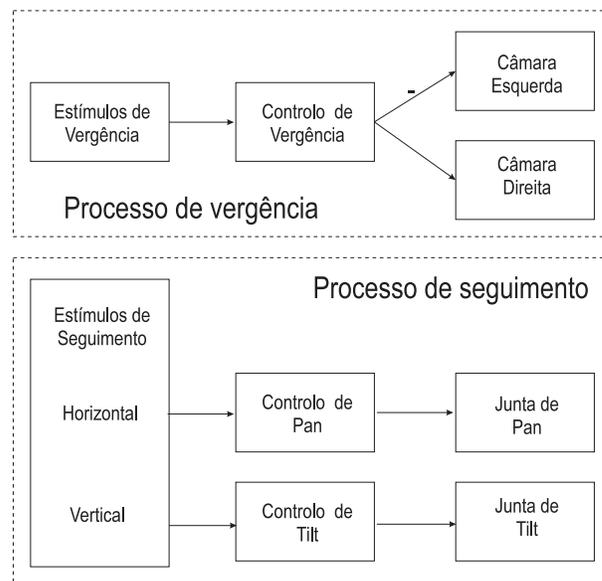


Fig. 6.6: Movimentos de vergência, seguimento horizontal e seguimento vertical são perfeitamente desacoplados.

No entanto, o processo de seguimento depende fortemente de uma apropriada segmentação do alvo, o que no nosso caso só é possível quando o alvo se encontra sob vergência ou no horóptero do observador. Este facto obriga a que o processo de vergência iniba o seguimento sempre que o alvo não esteja próximo de uma situação correcta de

vergência. Assim, a coordenação dos processos é efectuada de acordo com o diagrama da Figura 6.7.

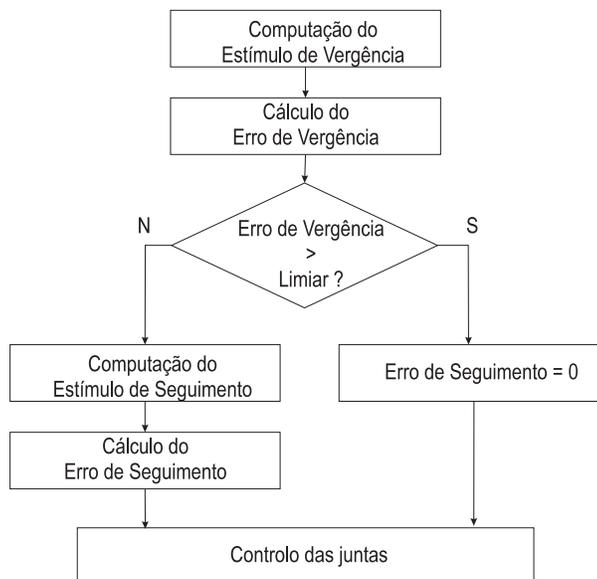


Fig. 6.7: Coordenação dos processos de vergência e de seguimento. O processo de seguimento só é activado quando há garantia que a segmentação do alvo produz resultados fiáveis.

A detecção da proximidade de uma situação vergência pode ser feita por inspecção do valor de erro de vergência. Sempre que este valor for superior a um determinado limiar então o processo de seguimento é inibido. O processo de vergência passa a funcionar isoladamente até que esse erro se torne suficientemente baixo.

7. Resultados

Neste capítulo apresentam-se resultados de experiências efectuadas com o sistema de seguimento de alvos móveis desenvolvido nesta tese. Relembramos que alguns resultados relacionados apenas com o movimento de vergência foram apresentados no Capítulo 4. Aqui considera-se o sistema completo, composto pela integração e coordenação dos movimentos de vergência com os movimentos de seguimento direccional.

A base experimental utilizada foi a cabeça robótica Medusa ligada a um computador PC Pentium 200 que adquire as imagens provenientes das duas câmaras a preto e branco montadas na cabeça robótica, através de uma placa IC-PCI (*Imaging Technology*). Também o comando dos motores das juntas da cabeça robótica e a implementação dos algoritmos desenvolvidos, incluindo todo o processamento de imagem e controlo do ciclo visual, são efectuados pelo mesmo computador.

Na Figura 7.1 apresenta-se um diagrama com as ligações entre o equipamento utilizado.

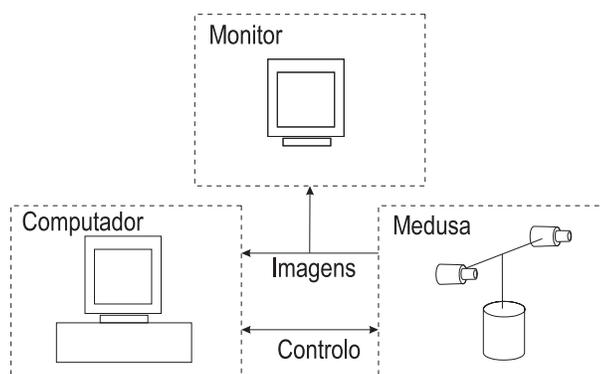


Fig. 7.1: Base experimental utilizada.

Em todas as experiências apresentadas neste capítulo, os algoritmos do ciclo de controlo visual são executados com uma frequência de 12.5 Hz. Na tabela seguinte apresentam-

-se os valores dos tempos de processamento dispendidos nas diversas fases de um ciclo de controlo.

Rotina de Processamento	Tempo de Cálculo (ms)
Aquisição de imagem	7.07
Mapeamento Log-Polar	4.93
Cálculo de disparidade (17 canais)	5.39
Segmentação do alvo	23.98
Estimação de posição e velocidade	6.99
Total	48.36

Para ilustrar o funcionamento, fiabilidade e generalidade do sistema proposto, as experiências foram divididas em três grupos principais:

- experiências em cadeia aberta, onde podemos observar os estímulos medidos pelo sistema apenas em função do movimento do alvo;
- experiências controladas em cadeia fechada, onde o movimento do alvo é conhecido;
- experiências em cadeia fechada em que o movimento do alvo não é conhecido.

Em todas estas experiências são utilizadas imagens log-polar com 64×32 *pixels*, correspondendo a uma redução de 1:128 na quantidade dos dados a processar, relativamente às imagens cartesianas completas (512×512 *pixels*).

7.1 Experiências em cadeia aberta

Um primeiro conjunto de experiências foi efectuado para avaliar qualitativa e quantitativamente o desempenho dos algoritmos de visão desenvolvidos, nomeadamente a estimação da posição e da velocidade do alvo no plano de imagem. Nestas experiências as câmaras mantêm-se estáticas e vergidas num ponto do espaço, enquanto se registam os valores das estimativas obtidas relativamente ao movimento de alvos que são feitos deslocar pelo

horóptero¹. Foram utilizados objectos de formas e padrões de textura variados, colocados a diversas distâncias do observador. Movimentos com velocidade constante foram preferencialmente utilizados para facilitar a avaliação da qualidade das estimativas obtidas.

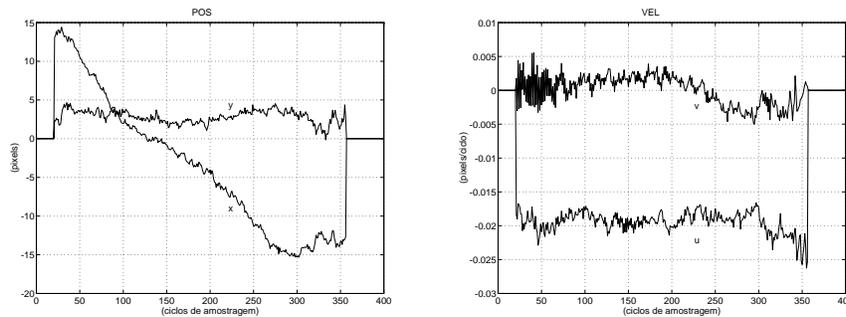


Fig. 7.2: Posição e velocidade estimadas para um alvo com velocidade constante.

Na Figura 7.2 apresentam-se as medidas de posição e de velocidade no plano de imagem de um alvo que se desloca horizontalmente a velocidade constante, a cerca de um metro em frente das câmaras. Embora de um ponto de vista macroscópico os gráficos revelem uma evolução coerente com o movimento do alvo, pode-se observar que localmente quer a posição quer a velocidade apresentam elevados níveis de ruído. Imperfeições e oscilações do movimento do alvo² assim como limitações próprias dos algoritmos utilizados (principalmente a sensibilidade do processo de segmentação) são causas importantes para este facto.

Ao longo destas experiências observou-se que, de um modo geral, as estimativas de velocidade revelavam melhores características para o controlo do seguimento do que as estimativas de posição. Embora este facto não seja óbvio a partir da Figura 7.2, pode ser observado na Figura 7.3.

Aqui comparam-se estimativas da posição do alvo através da medida de posição (1) e da integração numérica da medida de velocidade (2). É evidente um melhor comportamento das medidas de velocidade quer em precisão quer em imunidade ao ruído.

Através de experiências efectuadas em cadeia fechada comprovou-se que a utilização

¹só no horóptero é possível segmentar os pontos que pertencem ao alvo e proceder à estimação do movimento (ver Capítulo 5).

²a geração de velocidades constantes para o alvo foi feita de forma pouco precisa.

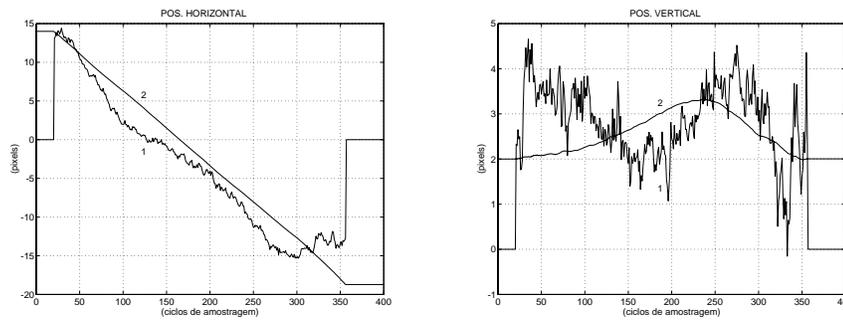


Fig. 7.3: Comparação entre estimativas de posição obtidas com medidas de posição (1) e velocidade (2). A estas últimas foi adicionado um valor constante para facilitar a comparação.

predominante de medidas de velocidade no controlo do processo de seguimento torna o seu desempenho bastante mais preciso e suave. No entanto, como resultam de uma integração numérica não permitem compensar erros iniciais de posição do alvo, pelo que é também necessário introduzir alguma informação de posição. Relembremos que o sinal de controlo resulta de uma média ponderada dos estímulos de posição e de velocidade (ver Capítulo 5). Pelas razões anteriores, é dado um maior factor de ponderação para as medidas de velocidade.

7.2 Experiências com movimento conhecido

Um segundo conjunto de experiências ilustra o desempenho do sistema de seguimento em alvos cujo movimento é conhecido. Os alvos são colocados inicialmente ao longo da direcção de observação. Após uma primeira fase em que ocorre apenas vergência, é dado início ao movimento do objecto.

Para cada uma das experiências procedeu-se ao registo das posições dos ângulos de cada junta da cabeça robótica. Através das relações cinemáticas do sistema (ver Capítulo 2), é possível determinar a posição do ponto de fixação no espaço e comparar com as trajectórias conhecidas para o alvo. É de referir que quer os movimentos do alvo quer os parâmetros de calibração do sistema robótico são conhecidos apenas qualitativamente. Assim, também os resultados obtidos só poderão ser avaliados qualitativamente.

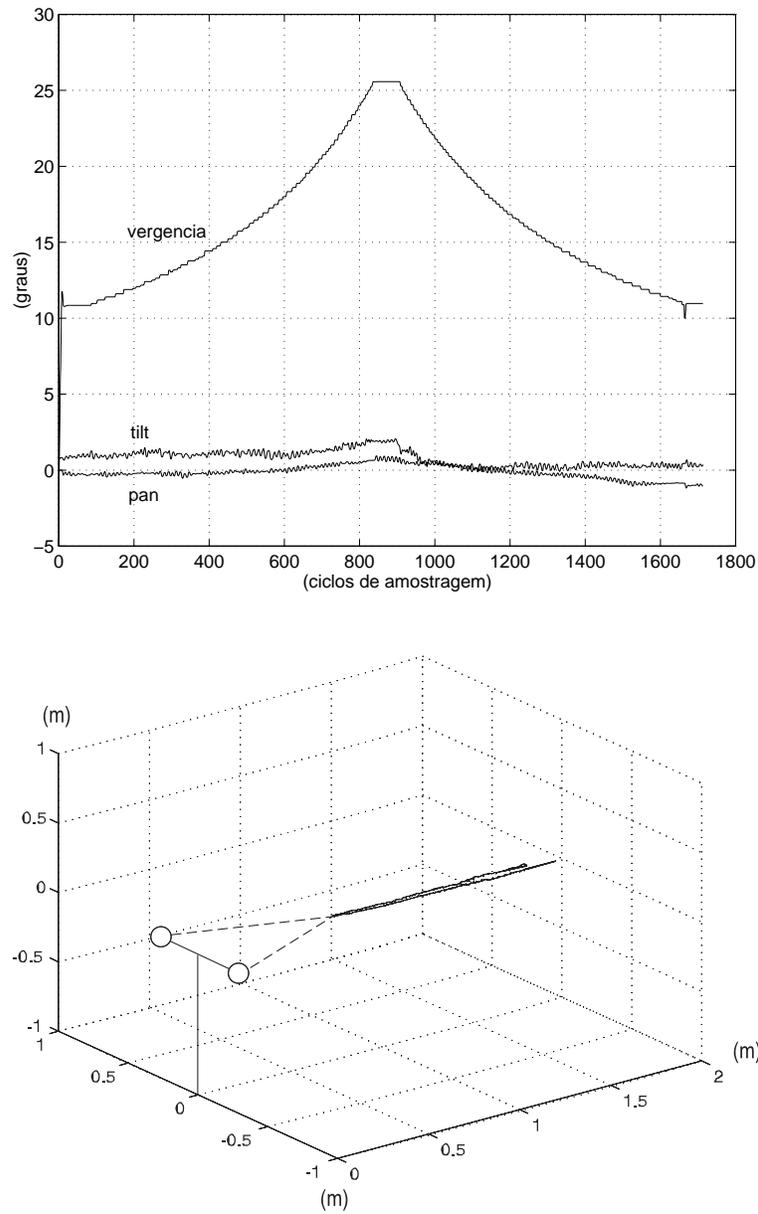


Fig. 7.4: Movimento em profundidade. Evolução temporal da posição angular das juntas (em cima) e trajectória reconstruída para o ponto de fixação (em baixo).

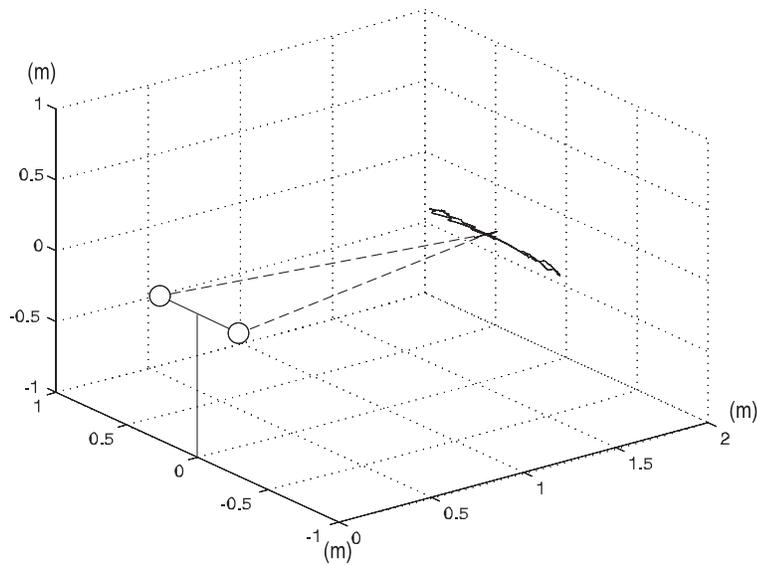
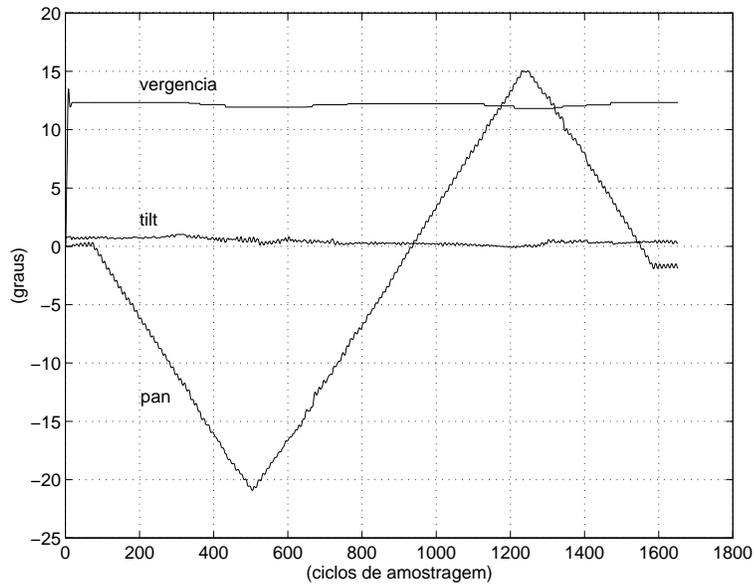


Fig. 7.5: Movimento fronto-paralelo. Evolução temporal da posição angular das juntas (em cima) e trajetória reconstruída para o ponto de fixação (em baixo).

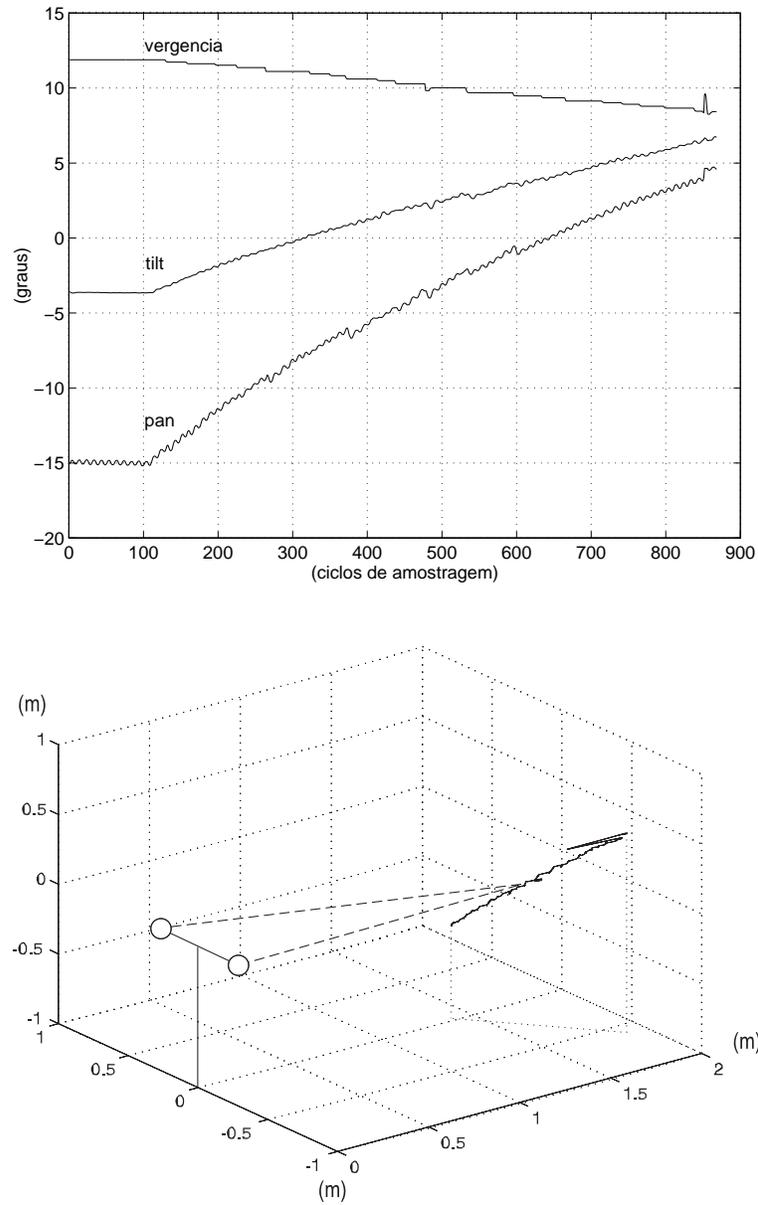


Fig. 7.6: Movimento retilíneo composto. Evolução temporal da posição angular das juntas (em cima) e trajectória reconstruída para o ponto de fixação (em baixo).

As Figuras 7.4, 7.5 e 7.6, ilustram experiências efectuadas com movimentos rectilíneos e de velocidade constante. Mostram-se os ângulos de vergência, de *tilt* e de *pan* e a trajectória reconstruída do ponto de fixação no espaço.

O primeiro caso refere-se a um movimento em profundidade. Como seria de esperar, apenas o ângulo de vergência apresenta variações significativas. A trajectória reconstruída para o ponto de fixação revela uma boa aproximação ao caso rectilíneo.

Na segunda experiência, o objecto move-se horizontalmente e numa direcção perpendicular à direcção de observação inicial. Novamente, a trajectória reconstruída permite confirmar a precisão do método. A maior parte do movimento é efectuado pela junta de *pan* mas também um pouco pela vergência. Os movimentos de *tilt* observados na figura devem-se principalmente a perturbações no movimento do alvo e a ruído no processo de estimação.

A terceira experiência ilustra um caso em que todos os graus de liberdade da cabeça robótica colaboram de forma significativa para o movimento global. O movimento imprimido ao alvo é de velocidade constante, à semelhança dos casos anteriores, mas a direcção do deslocamento é diagonal aos eixos do sistema de coordenadas adoptado na representação gráfica. Mais uma vez, a trajectória reconstruída tem uma forma aproximadamente rectilínea, à semelhança do movimento do alvo.

Na Figura 7.7 apresenta-se uma última experiência neste conjunto que ilustra o seguimento de um alvo que se move segundo uma trajectória circular com o centro próximo do eixo de rotação da junta de *pan*. Este é um movimento puro em azimute pelo que só induz movimentos significativos na junta de *pan*.

Todas estas experiências evidenciam uma boa capacidade do sistema em efectuar um seguimento bastante preciso de deslocamentos simples e controlados do alvo. No próximo conjunto de experiências isso também será verificado para movimentos desconhecidos e relativamente complexos.

7.3 Experiências com movimento desconhecido

Nesta secção apresenta-se um conjunto de experiências que, ao contrário das anteriores, não admitem um modelo de movimento para o alvo. Na Figura 7.8, apresenta-se a re-

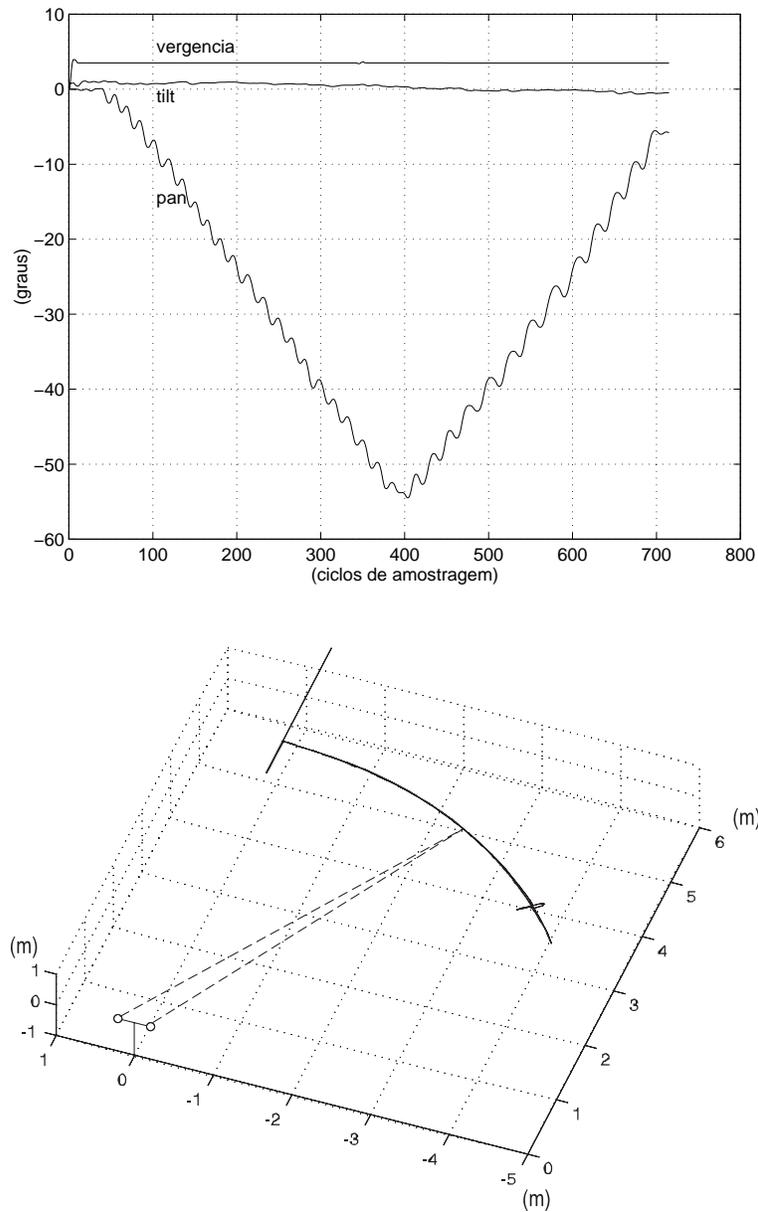


Fig. 7.7: Movimento circular. Evolução temporal da posição angular das juntas (em cima) e trajetória reconstruída para o ponto de fixação (em baixo).

construção da trajectória do ponto de fixação no seguimento de uma pessoa que se move arbitrariamente no interior de uma sala.

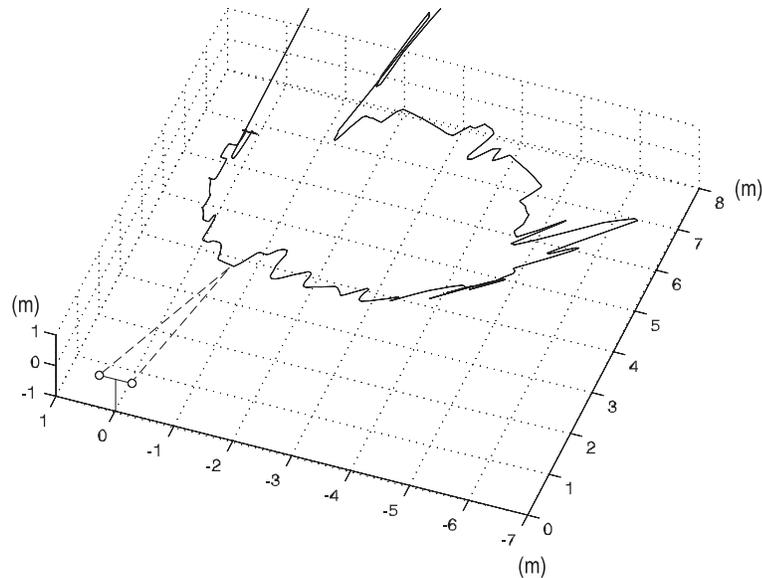


Fig. 7.8: Movimento livre. Trajectória reconstruída para o ponto de fixação.

Uma vez que não é conhecida a trajectória real do alvo, a reconstrução da trajectória do ponto de fixação não nos permite avaliar o desempenho do sistema. Por isso, a apresentação dos resultados destas experiências é feita através das imagens obtidas no decorrer dos processos de seguimento. As imagens são representadas na geometria adoptada para a sua utilização interna (log-polar). Um bom seguimento será verificado se o alvo se mantiver próximo do centro das imagens da sequência.

Para ilustrar a generalidade do sistema face à forma ou movimento dos alvos, são apresentadas imagens resultantes de três experiências efectuadas com diferentes objectos e com diferentes tipos de movimento.

A primeira sequência é apresentada na Figura 7.9 e consiste no seguimento de uma mão que se move arbitrariamente no interior de uma sala. Inicialmente a cabeça robótica Medusa mantém fixação na porta ao fundo da sala (imagens 1 e 2 da sequência). A partir do instante em que a mão se aproxima da direcção de observação (imagem 3) o processo de vergência efectua a fixação no alvo. Em seguida, dá-se o seguimento do movimento da mão pelo campo visual. É de realçar a boa precisão no seguimento (a mão mantém-se

sempre muito próximo do centro da imagem), apesar de existir rotação e escalamento da forma do alvo durante o seu deslocamento. Ainda relativamente a esta sequência, mostra-se na Figura 7.10 o resultado da segmentação do objecto para cada imagem. Note-se que apesar do complexo deslocamento do fundo induzido pelo movimento das câmaras, a mão é o único objecto segmentado.

A segunda sequência, ilustrada na Figura 7.11, refere-se a uma situação de seguimento do rosto de uma pessoa que se desloca pela mesma sala. Esta sequência permite observar o bom desempenho do sistema perante fortes escalamentos do alvo (comparar imagens entre 16 e 20) e variações da pose do alvo (comparar imagens 3 e 4). A alvo desloca-se predominantemente em direcções horizontais mas entre as imagens 12 e 17 pode-se observar um movimento vertical.

Por fim, a terceira sequência, apresentada na Figura 7.12, demonstra o comportamento do sistema perante uma maior proximidade entre objectos e oclusões. Inicialmente o sistema mantém fixação no objecto 1 (Pinguim) enquanto o objecto 2 (Pai Natal) se aproxima. Na terceira imagem da sequência o objecto 2 passa à frente do objecto 1. O sistema fixa o objecto 2, dado que este oclui o primeiro, e segue-o entre as imagens 3 e 13. A partir deste instante é o objecto 1 que se sobrepõe ao objecto 2. O sistema segue o objecto 1 até ao fim da sequência.

A última sequência apresentada é bastante complexa e com objectos relativamente pequenos. Mesmo assim, efectuou-se um seguimento suficientemente preciso dos alvos. De um modo geral, para objectos não demasiado pequenos e com movimentos não muito bruscos, o sistema de seguimento funciona com bastante fiabilidade.

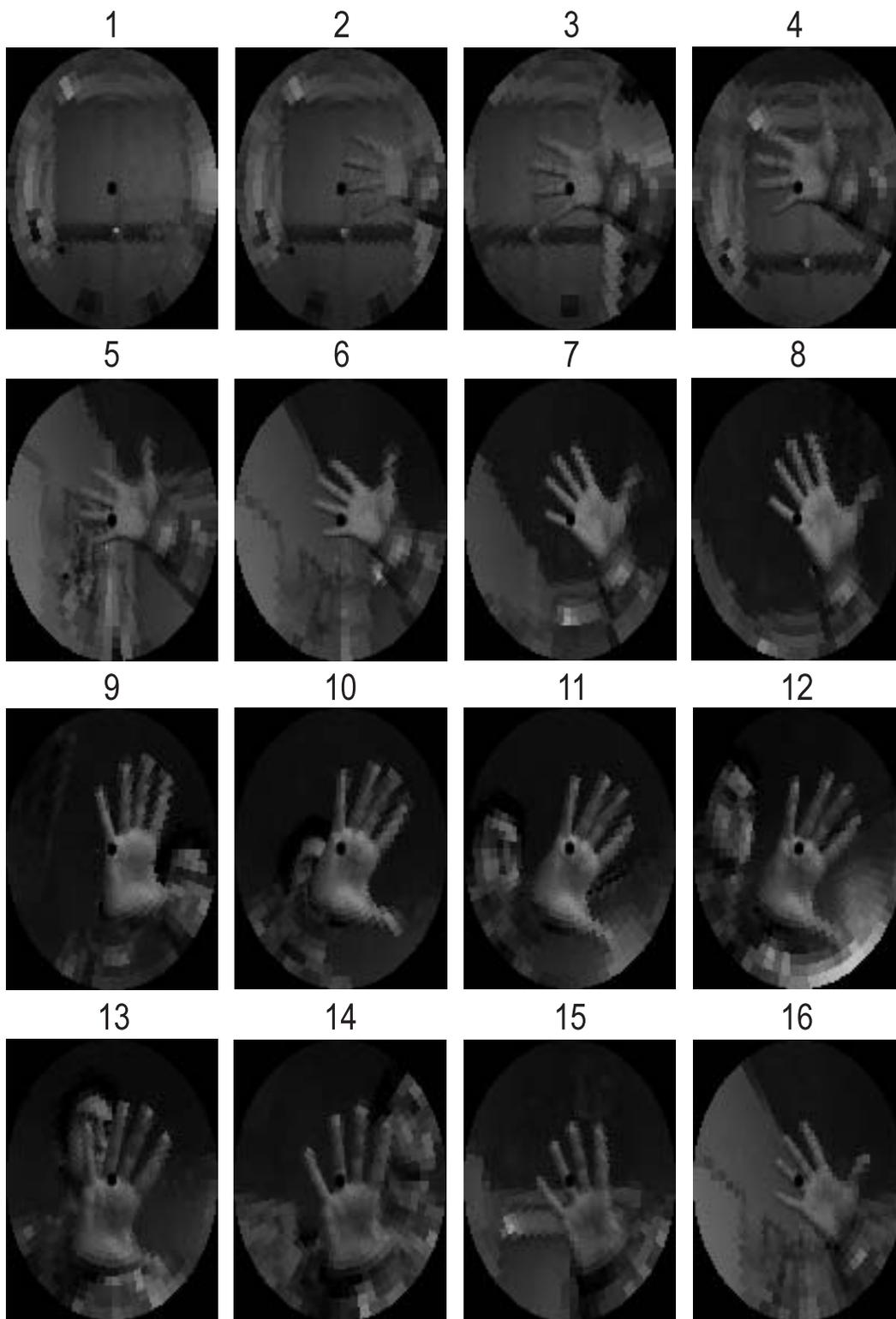


Fig. 7.9: Sequência 1. Imagens obtidas pela câmara esquerda.

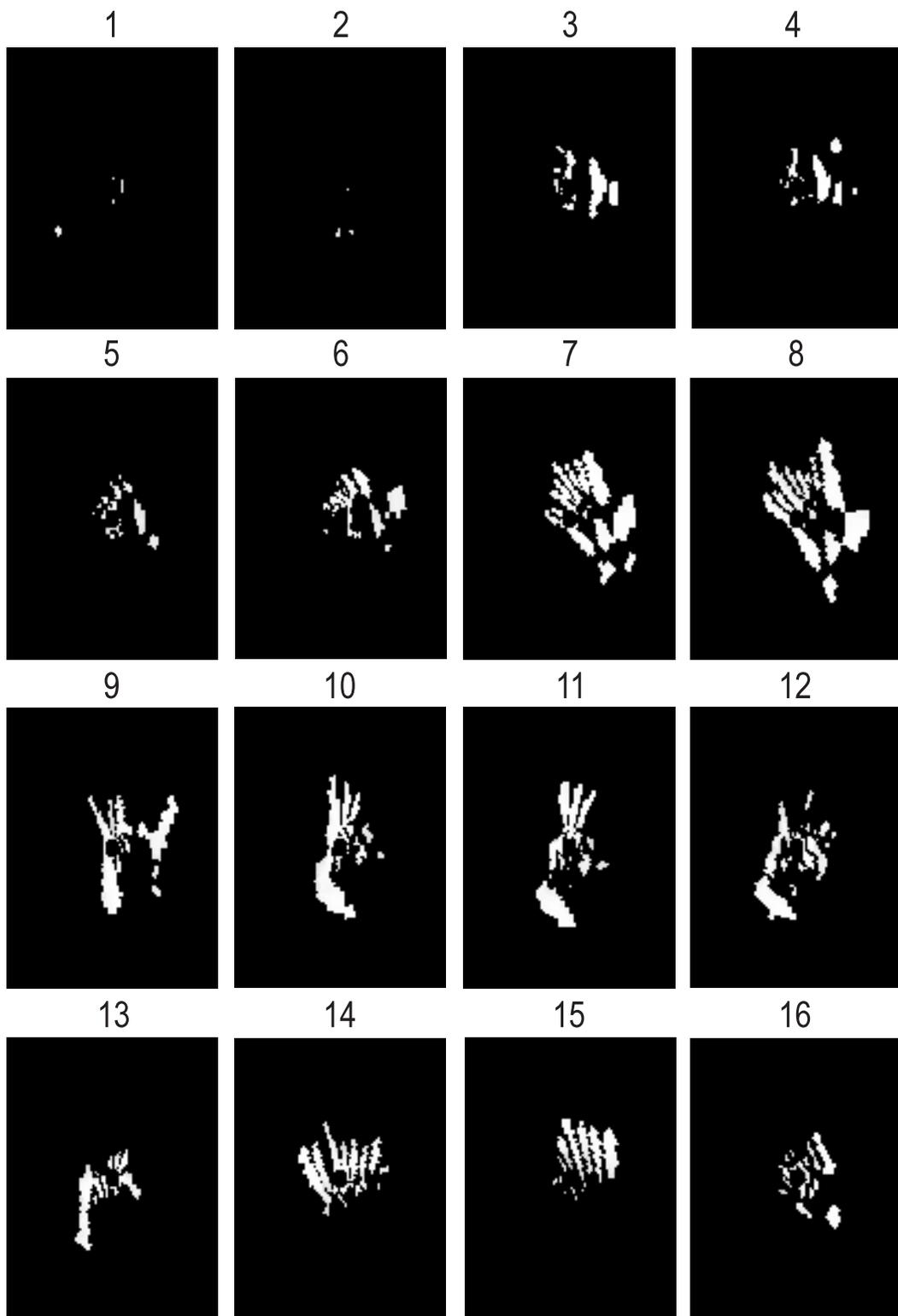


Fig. 7.10: Sequência 1. Resultados da segmentação do objecto.

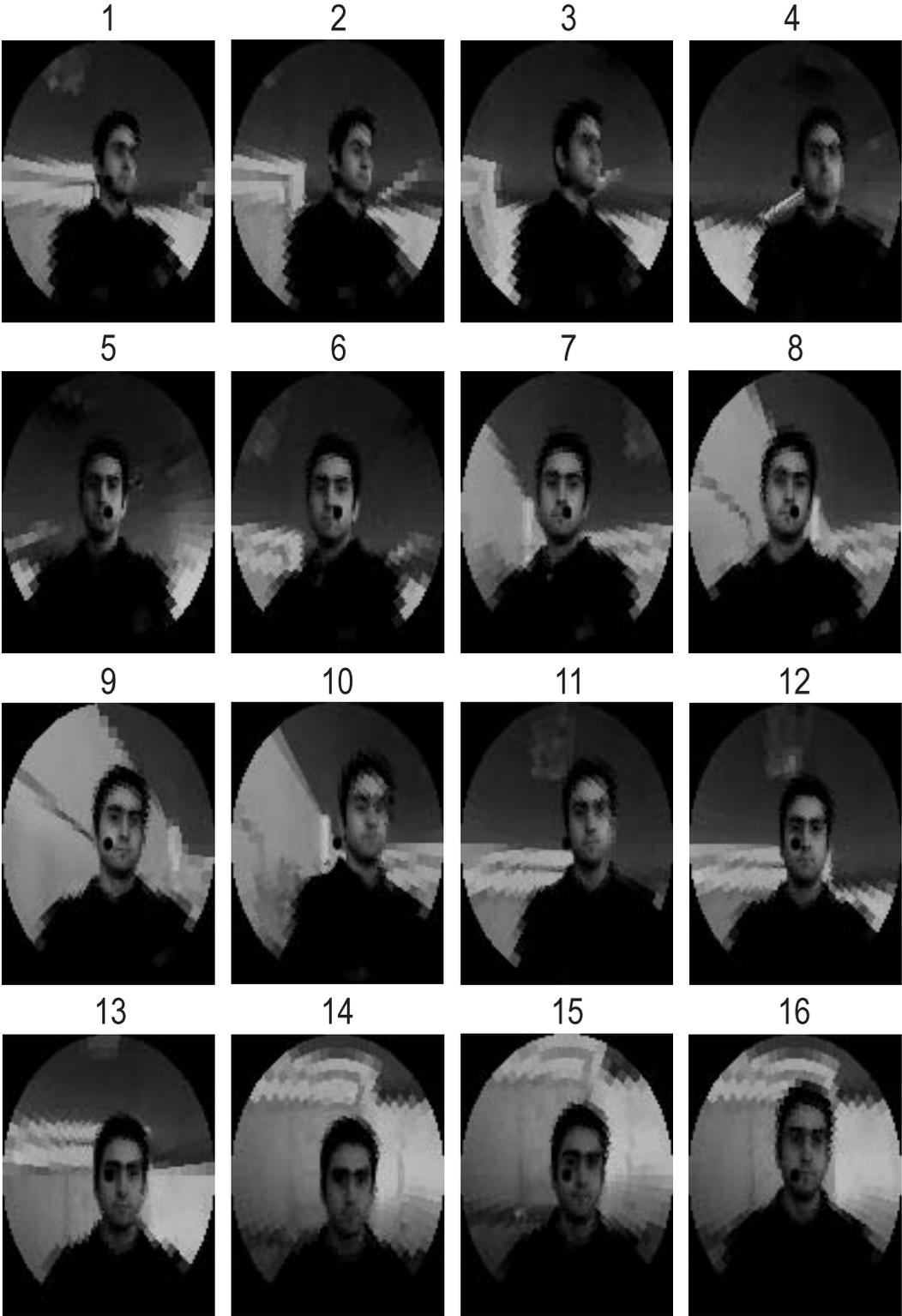
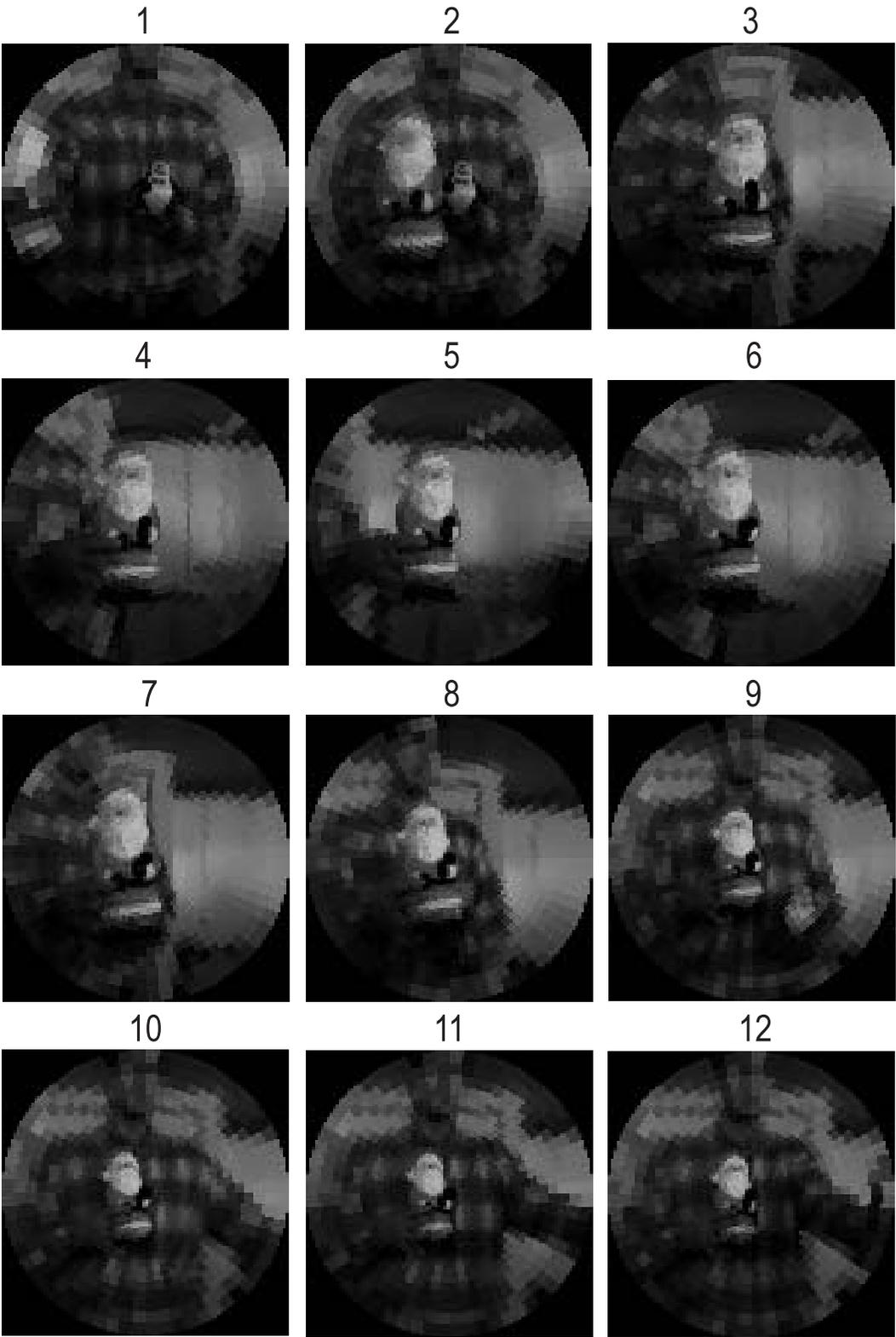




Fig. 7.11: Sequência 2. Imagens obtidas pela câmara esquerda.



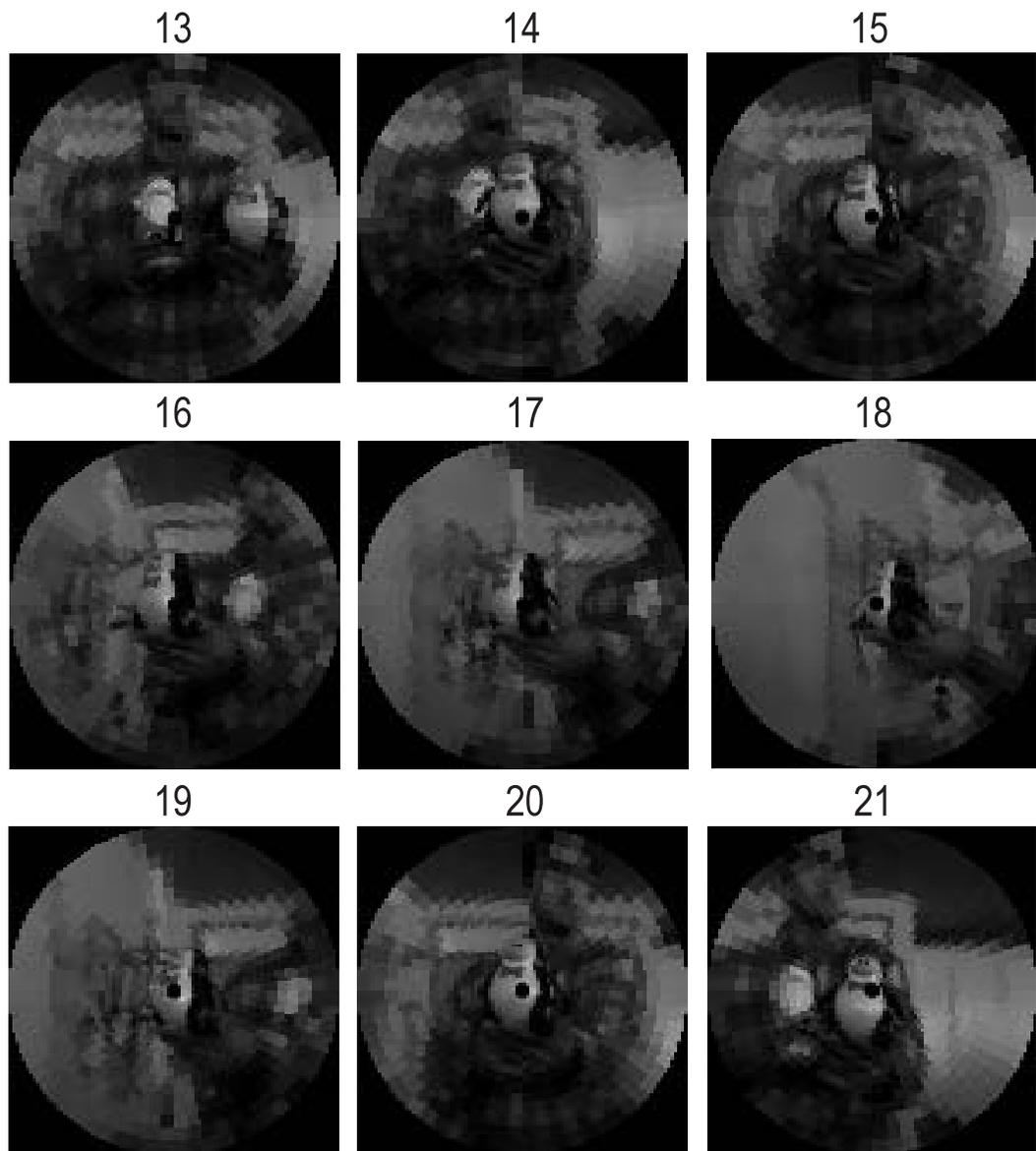


Fig. 7.12: Sequência 3. Imagens obtidas pela câmara esquerda.

8. Conclusões

Esta tese estuda um problema com grande importância no domínio da visão robótica: a fixação e seguimento de alvos móveis. Um sistema concreto é proposto e implementado na cabeça robótica Medusa. Uma das grandes preocupações evidenciadas no projecto deste sistema centra-se na generalidade de aplicação no que respeita à aparência visual dos alvos e ao seu movimento relativo ao observador.

A abordagem utilizada é motivada biologicamente através da análise do sistema visual humano e de outros animais, e enquadra-se no paradigmas da Visão Activa e da Visão Objectiva. Neste contexto, três direcções principais são tomadas: a representação de imagens através de uma geometria de resolução não uniforme, próxima da utilizada na retina humana; a utilização de informação visual específica, de rápida e fácil extracção, para controlar directamente os graus de liberdade do sistema robótico; a decomposição dos movimentos oculares motivada pelo que acontece no sistema de controlo oculomotor humano. Os resultados apresentados revelam a validade da abordagem tomada. Objectos de várias formas e com diversos tipos de movimento são seguidos em tempo-real e de uma forma robusta.

A utilização de imagens log-polar mostrou-se especialmente adequada na obtenção dos estímulos visuais utilizados no controlo do seguimento de alvos móveis. Como estas imagens dotam a zona central com maior resolução do que a periferia, e dado que é nesta zona que o alvo se mantém a maior parte do tempo, o movimento do alvo têm uma influência predominante nos estímulos obtidos. Assim, elementos distractores no espaço visual, tais como outros objectos móveis ou o deslocamento do fundo induzido pelo movimento das câmaras, não degradam de forma significativa o desempenho do sistema. Adicionalmente, a transformação log-polar proporciona imagens com dimensões

inferiores às imagens cartesianas, o que torna o seu processamento computacionalmente menos exigente, para igual campo visual e detalhe no centro das imagens.

A introdução de um conjunto de conceitos relacionado com resolução de uma imagem, e uma forma de representação de imagens baseada em transformações de coordenadas, permitiu analisar, com base no *Jacobiano* da transformação, a relação existente entre o processamento de imagem no domínio cartesiano e em domínios de resolução variante no espaço. Em particular a aplicação de algoritmos de correlação em imagens log-polar corresponde a um mecanismo atencional implícito que “pesa” zonas da imagem cartesiana correspondente em função da sua distância à origem.

Um dos factores determinantes para a generalidade do sistema de seguimento face a alterações de forma ou de movimento do alvo deve-se ao uso de pistas visuais de baixo nível nos processos de sensoriameto. As técnicas utilizadas para a extracção dessas pistas, tais como a correlação de imagens, a filtragem de disparidade nula, ou o cálculo do fluxo óptico, não pressupõem conhecimento prévio sobre a forma ou o movimento do alvo, sendo por isso utilizáveis para uma grande diversidade de situações. Para além disso, estas pistas podem ser obtidas através de algoritmos de processamento simples e rápidos, apesar de pouco precisos, permitindo o funcionamento do sistema em tempo-real. A menor precisão dos algoritmos é compensada pela integração das medidas numa arquitectura de controlo dinâmico em cadeia fechada. Este tipo de controlo proporciona ainda uma movimentação suave do sistema apesar do ruído contido no sinais estimados.

Através do estudo de sistemas biológicos podemos encontrar muitos indícios com potencial utilidade no projecto de sistemas artificiais. Neste trabalho, a decomposição dos movimentos oculares em componentes separáveis de vergência e de seguimento é motivada pelo sistema visual humano e revelou-se fundamental nos resultados obtidos. A característica de separabilidade simplifica a análise cinemática do sistema e facilita o projecto dos processos de controlo.

É também interessante notar que os processos de vergência e de seguimento cooperam e dependem entre si de forma muito directa. Por um lado, o processo de vergência compensa os movimentos de profundidade de alvos que se encontrem na direcção de observação. Por outro lado, o processo de seguimento corrige a direcção de observação para objectos que se encontrem sob vergência.

Outro aspecto importante no desempenho do sistema consiste no aproveitamento da fusão binocular proporcionada pela vergência. Nesta circunstância torna-se possível efectuar de uma forma simples a segmentação das zonas da imagem pertencentes ao alvo de interesse. Os algoritmos de estimação de posição e movimento do alvo beneficiam deste facto pois só necessitam de considerar estas zonas da imagem, o que pode reduzir consideravelmente o processamento efectuado.

8.1 Trabalho futuro

Desenvolvimentos futuros a este trabalho poderão tomar várias direcções. Em primeiro lugar, apesar dos bons resultados obtidos, o sistema de seguimento proposto nesta tese pode ainda ser aperfeiçoado. Certos aspectos próprios ao problema de seguimento que não foram aprofundados neste trabalho, podem ser relevantes no desempenho global do sistema. Como lidar com oclusões temporárias de uma das câmaras, com deformações sérias de perspectiva devidas a elevados ângulos de vergência ou com múltiplos objectos nas proximidades do alvo de interesse, são problemas que podem beneficiar da integração de outros tipos de informação, visual ou não visual. Estímulos de desfocagem de imagem, informação de côr e cálculo de correlação ou fluxo óptico a diversas escalas são pistas visuais de baixo nível e de fácil e rápida extracção que podem tornar mais robusto o desempenho do sistema.

Seria interessante considerar a utilização de graus de liberdade ópticos para ajuste do escalamento de objectos observados ou de sensores de aceleração para compensar o movimento próprio, de modo a avaliar se o investimento nestes dispositivos dispendiosos é justificado pelas melhorias no desempenho do sistema.

Seria também interessante explorar a influência da geometria da zona atencional implícita, proporcionada pela utilização de imagens de resolução não uniforme. Por exemplo, a sua dimensão poderia ser automaticamente ajustada em função da dimensão relativa dos alvos observados. A utilização de imagens com resolução Gaussiana permitiria a configuração da dimensão do foco atencional através do ajuste da variância da distribuição e evitaria a singularidade apresentada na origem pela geometria log-polar. O uso de superfícies Gaussianas para considerar de diferente forma os objectos presentes no centro da

imagem e os objectos da periferia é sugerida em [23], num contexto semelhante.

Para além dos problemas intrínsecos ao comportamento de fixação de atenção, há outros aspectos que podem ser bastante úteis para a autonomia o processo de seguimento. Como seleccionar inicialmente o alvo de fixação ou como alterar o alvo de interesse durante o seguimento são assuntos igualmente importantes no desempenho de um observador artificial e que dependem de uma componente motivacional própria. A análise dos sistemas de visão biológicos a um nível cognitivo mais elevado pode sugerir metodologias para a abordagem destes problemas.

Outro aspecto interessante seria considerar a integração do sistema de seguimento de alvos móveis em tarefas de navegação, manipulação ou reconhecimento. Por exemplo [26] estabelece teoricamente as vantagens da fixação ocular em objectos do espaço para efectuar estimação do movimento tridimensional do objecto, recuperação do movimento próprio ou estimação do tempo para contacto. Também no contexto do reconhecimento de objectos, a fixação visual pode trazer vantagens importantes [18].

A. Transformação log-polar

A transformação log-polar é definida como:

$$\begin{cases} \mathbf{T} : R^2 \setminus \{0\} \rightarrow R_0^+ \times R_0^+ \\ (\xi, \eta) = \mathbf{T}(x, y) = \left(\log_k \frac{\sqrt{x^2+y^2}}{\rho_{\min}}, \alpha \arctan\left(\frac{y}{x}\right) \right) \end{cases} \quad (\text{A.1})$$

As constantes k , ρ_{\min} e α são parâmetros da transformação e a função $\arctan(\cdot)$ é definida como tendo resultados no intervalo $[0, 2\pi[$.

A matriz Jacobiana desta transformação é dada por:

$$\mathbf{J}(x, y) = \begin{bmatrix} \frac{\partial \xi}{\partial x} & \frac{\partial \xi}{\partial y} \\ \frac{\partial \eta}{\partial x} & \frac{\partial \eta}{\partial y} \end{bmatrix} = \begin{bmatrix} \frac{1}{\log k} \frac{x}{x^2+y^2} & \frac{1}{\log k} \frac{y}{x^2+y^2} \\ -\alpha \frac{y}{x^2+y^2} & \alpha \frac{x}{x^2+y^2} \end{bmatrix} \quad (\text{A.2})$$

O determinante da matriz Jacobiana (o *Jacobiano*) é:

$$|\mathbf{J}(x, y)| = \frac{\alpha}{\log k} \frac{1}{x^2 + y^2} \quad (\text{A.3})$$

A transformação inversa da transformação log-polar é definida como:

$$\begin{cases} \mathbf{T}^{-1} : R_0^+ \times R_0^+ \rightarrow R^2 \setminus \{0\} \\ (x, y) = \mathbf{T}^{-1}(\xi, \eta) = \left(\rho_{\min} k^\xi \cos \frac{\eta}{\alpha}, \rho_{\min} k^\xi \sin \frac{\eta}{\alpha} \right) \end{cases} \quad (\text{A.4})$$

Tem como matriz Jacobiana:

$$\mathbf{J}^{-1}(\xi, \eta) = \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial x}{\partial \eta} \\ \frac{\partial y}{\partial \xi} & \frac{\partial y}{\partial \eta} \end{bmatrix} = \begin{bmatrix} \log k \rho_{\min} k^\xi \cos \frac{\eta}{\alpha} & -\frac{\rho_{\min}}{\alpha} k^\xi \sin \frac{\eta}{\alpha} \\ \log k \rho_{\min} k^\xi \sin \frac{\eta}{\alpha} & \frac{\rho_{\min}}{\alpha} k^\xi \cos \frac{\eta}{\alpha} \end{bmatrix} \quad (\text{A.5})$$

O *Jacobiano* vem dado por:

$$|\mathbf{J}^{-1}(\xi, \eta)| = \frac{\log k}{\alpha} \left(\rho_{\min} k^\xi \right)^2 \quad (\text{A.6})$$

B. Dimensionamento dos controladores de seguimento

Os controladores P.I.D. na sua forma discreta são representados pela seguinte função de transferência:

$$D(z) = K_p \left(1 + \frac{T}{T_i} \frac{z}{z-1} + \frac{T_d}{T} \frac{z-1}{z} \right) \quad (\text{B.1})$$

Os termos K_p , T_i e T_d na expressão anterior designam-se por “ganho proporcional”, “tempo integral” e “tempo derivativo”, respectivamente, e T é o período de amostragem do ciclo de controlo.

Em especial nos problemas de controlo de motores, a experiência acumulada em projecto de controladores dinâmicos sugere a aplicação de controladores P.I. (tempo derivativo nulo) em controlo de velocidade e de controladores P.D. (tempo integral infinito) em controlo de posição, para a obtenção de boas características dinâmicas na resposta do sistema em malha fechada.

Um controlador P.I. discreto tem a seguinte função de transferência:

$$D(z) = K_p \left(1 + \frac{T}{T_i} \frac{z}{z-1} \right) \quad (\text{B.2})$$

Um controlador P.D. discreto tem a seguinte função de transferência:

$$D(z) = K_p \left(1 + \frac{T_d}{T} \frac{z-1}{z} \right) \quad (\text{B.3})$$

O comando das juntas é efectuado em velocidade dado que proporciona respostas mais suaves do que o comando em posição. As Equações (5.6) referentes ao controlo de posição, são reescritas sob a forma de funções de transferência, tendo em conta este facto:

$$\left\{ \begin{array}{l} \frac{X(z)}{\Delta W_{pan}(z)} = \frac{T}{\gamma_x} \frac{1}{z(z-1)} \\ \frac{Y(z)}{\Delta W_{tilt}(z)} = \frac{T}{\gamma_y} \frac{1}{z(z-1)} \end{array} \right. \quad \text{e} \quad \left\{ \begin{array}{l} \frac{V_x(z)}{\Delta W_{pan}(z)} = \frac{1}{\gamma_u} \frac{1}{z^2} \\ \frac{V_y(z)}{\Delta W_{tilt}(z)} = \frac{1}{\gamma_v} \frac{1}{z^2} \end{array} \right.$$

onde $X(z)$, $Y(z)$, $V_x(z)$, $V_y(z)$, $\Delta W_{pan}(z)$ e $\Delta W_{tilt}(z)$ são as transformadas Z dos sinais $x(t)$, $y(t)$, $v_x(t)$, $v_y(t)$, $\delta w_{pan}(t)$ e $\delta w_{tilt}(t)$, respectivamente. Os ganhos γ_x , γ_y , γ_u e γ_v pode ser considerados unitários, sem perda de generalidade, para simplificar o dimensionamento.

B.1 Controlo de posição

Para efectuar o controlo de posição do sistema introduz-se de um controlador P.D. (Equação (B.3)) em série com a malha de retroacção de posição. Obtém-se o seguinte ganho de malha¹:

$$L_\theta(z) = D_\theta(z) G_\theta(z) = K_p T_d \frac{z \left(\frac{T+T_d}{T_d} \right) - 1}{z^2 (z - 1)}$$

De modo a dotar o sistema com boas propriedades dinâmicas em malha fechada devemos proceder a uma escolha adequada para a localização dos polos, à custa do ajuste do valor do *zero*, $(T + T_d)/T_d$, e do ganho K_p . A opção tomada considerou uma escolha de polos sobre o eixo real, de modo a obter respostas sem sobreelevação e, o mais próximo possível da origem, para aumentar a rapidez do sistema, o que corresponde à colocação dos polos em localizações idênticas sobre o eixo real. Este dimensionamento conduz ao seguintes parâmetros:

$$K_p = \frac{8}{27 T} \quad \text{e} \quad T_d = \frac{T}{8}$$

e a uma localização dos polos em:

$$p = \frac{1}{3}$$

B.2 Controlo de velocidade

A aplicação de um controlador P.I. (Equação (B.2)) para o controlo de velocidade conduz ao seguinte ganho de malha:

$$L_\omega(z) = D_\omega(z) G_\omega(z) = K_p \frac{z \left(\frac{T+T_i}{T_i} \right) - 1}{z^2 (z - 1)}$$

¹produto das funções de transferência existentes num ciclo de controlo.

Pelas mesmas razões invocadas para o controlo de posição, colocam-se os polos em malha fechada em iguais localizações sobre o eixo real. Assim, ficamos com:

$$K_p = \frac{1}{27} \quad \text{e} \quad T_i = \frac{T}{8}$$

e a mesma localização dos polos do que para o controlo de posição.

Referências

- [1] Y. Aloimonos. Purposive and qualitative active vision. In *Proc. of the 10th. IEEE Int. Conference on Pattern Recognition*, Atlantic City, NJ - USA, June 1990.
- [2] Y. Aloimonos. Introduction: Active vision revisited. In Yiannis Aloimonos, editor, *Active Perception*. Lawrence Erlbaum Associates, 1993.
- [3] C. Andersen and H. Christensen. Integration of visual processes for control of fixation. In *Proc. of the 4th Int. Symposium on Intelligent Robotic Systems*, Lisbon, Portugal, July 1996.
- [4] T. Anderson and M. Donath. Animal behavior as a paradigm for developing robot autonomy. In P. Maes, editor, *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back*, pages 145–168. MIT/Elsevier, 1991.
- [5] A. Argyros, M. Lourakis, P. Trahanias, and S. Orphanoudakis. Qualitative detection of 3d motion discontinuities. In *Proc. of the 1996 IEEE/RJS International Conference on Intelligent Robots and Systems*, pages 1630–1637, Osaka, Japan, November 1996. IEEE Computer Society Press.
- [6] R. Arkin. Integrating behavioral, perceptual and world knowledge in reactive navigation. In P. Maes, editor, *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back*, pages 105–122. MIT/Elsevier, 1990.
- [7] P. Aschwanden and W. Guggenbühl. Experimental results from a comparative study on correlation-type registration algorithms. In W. Förstner and St. Ruwiedel, editors, *Robust Computer Vision*. Wichmann, 1992.

- [8] R. Bajcsy. Active perception. *Proc. of the IEEE*, 76(8):996–1005, August 1988.
- [9] D. Ballard. Animate vision. *Artificial Intelligence*, 48:57–86, 1991.
- [10] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–78, February 1994.
- [11] R. Beer, H. Chiel, and L. Sterling. A biological perspective on autonomous agent design. In P. Maes, editor, *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back*, pages 169–186. MIT/Elsevier, 1990.
- [12] A. Bernardino and J. Santos-Victor. Correlation based vergence control using log-polar images. In *Proc. of the 4th Int. Symposium on Intelligent Robotic Systems*, Lisbon, Portugal, July 1996.
- [13] A. Bernardino and J. Santos-Victor. Sensor geometry for dynamic vergence: characterization and performance analysis. In *Proc. of the Workshop on Performance Characteristics of Vision Algorithms*, Cambridge, UK, April 1996.
- [14] A. Bernardino and J. Santos-Victor. Vergence control for robotic heads using log-polar images. In *Proc. of the 1996 IEEE/RJS International Conference on Intelligent Robots and Systems*, pages 1264–1271, Osaka, Japan, November 1996. IEEE Computer Society Press.
- [15] A. Bernardino, C. Silva, J. Santos-Victor, and C. Pinto-Ferreira. Behaviour based oculomotor control architecture for stereo heads. In *Proc. of the 3rd Int. Symposium on Intelligent Robotic Systems*, Pisa, Italy, July 1995.
- [16] K. Brenner, G. Lohman, H. Buchkremer, A. Jensen, E. Rasmussen, H. Rajbenbach, and J. Huignard. New architectures for optical processing in industrial applications. In *Proc. Esprit Day, ECCV*, Santa Margherita Ligure, Italy, May 1992.
- [17] R. Brooks. A robust layered control system for a mobile robot. *IEEE J. Robotics and Automation*, pages 14–23, April 1986.
- [18] K. Brunnström, J. Eklundh, and T. Uhlin. Active fixation for scene exploration. *Int. Journal of Computer Vision*, 17(2):137–162, February 1996.

- [19] C. Capurro, F. Panerai, and G. Sandini. Dynamic vergence. In *Proc. of the 1996 IEEE/RJS International Conference on Intelligent Robots and Systems*, pages 1241–1247, Osaka, Japan, November 1996. IEEE Computer Society Press.
- [20] C. Capurro, F. Panerai, and G. Sandini. Fusion based vergence and tracking. In *Proc. of the 4th Int. Symposium on Intelligent Robotic Systems*, Lisbon, Portugal, July 1996.
- [21] R. Carpenter. *Movements of the eyes*. Pion, London, 1988.
- [22] H. Christensen, K. Bowyer, and H. Bunke, editors. *Active Robot Vision : camera heads, model based navigation and reactive control*, volume 6. World Scientific, 1994.
- [23] D. Coombs. *Real-time Gaze Holding in Binocular Robot Vision*. PhD thesis, University of Rochester, Rochester, New York, October 1991.
- [24] D. Coombs and C. Brown. Real-time binocular smooth pursuit. *Int. Journal of Computer Vision*, 11(2):147–164, October 1993.
- [25] G. Dagnelie and R. Massof. Toward an artificial eye. *IEEE Spectrum*, 1:21–29, May 1996.
- [26] C. Fermüller and Y. Aloimonos. The role of fixation in visual motion analysis. *Int. Journal of Computer Vision*, 11(2):165–186, October 1993.
- [27] F. Ferrari, J. Nielsen, P. Questa, and G. Sandini. Space variant sensing for personal communication and remote monitoring. In *Proc. of the EU-HCM Smart Workshop*, Lisbon, Portugal, April 1995.
- [28] D. Fleet and A. Jepson. Computation of component image velocity from local phase information. *Int. Journal of Computer Vision*, 5(1):77–104, 1990.
- [29] D. Fleet, A. Jepson, and M. Jenkin. Image matching using the windowed fourier phase. *CVGIP: Image Understanding*, 53(2):198–210, March 1991.
- [30] G. Franklin, J. Powell, and M. Workman. *Digital Control of Dynamic Systems*. Addison-Wesley, 1990.

- [31] D. Gennery. Visual tracking of known three-dimensional objects. *International Journal of Computer Vision*, 7(3):243–270, April 1992.
- [32] N. Griswald, J. Lee, and C. Weiman. Binocular fusion revisited utilizing a log-polar tessellation. *Computer Vision and Image Processing*, pages 421–457, 1992.
- [33] B. Horn. *Robot Vision*. MIT Press, McGraw Hill, 1986.
- [34] B. Horn and B. Shunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.
- [35] A. Jain. *Fundamentals of digital image processing*. Prentice Hall, New Jersey, 1989.
- [36] D. Koller, K. Daniilidis, and H. Nagel. Model-based object tracking in monocular image sequences of road traffic scenes. *International Journal of Computer Vision*, 10(3):257–281, June 1993.
- [37] E. Krotkov and R. Bajcsy. Active vision for reliable ranging: Cooperating focus, stereo and vergence. *Int. Journal of Computer Vision*, 11(2):187–203, October 1993.
- [38] M. Land. Similarities in the visual behavior of arthropods and men. In Michael Gazzaniga and Collin Blakemore, editors, *Handbook of Psychobiology*. Academic Press, 1975.
- [39] D. Lowe. Robust model-based motion tracking through the integration of search and estimation. *International Journal of Computer Vision*, 8(2):113–122, August 1992.
- [40] K. Ludwig, H. Neumann, and B. Neumann. Robust estimation of local stereoscopic depth. In W. Förstner and St. Ruwiedel, editors, *Robust Computer Vision*. Wichmann, 1992.
- [41] P. Maes. Situated agents can have goals. In P. Maes, editor, *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back*, pages 49–70. MIT/Elsevier, 1991.

- [42] R. Milanese. *Detecting salient regions in an image: from biological evidence to computer implementation*. PhD thesis, University of Geneva, Geneva, Switzerland, December 1993.
- [43] D. Murray, K. Bradshaw, P. McLauchlan, and I. Reid. Driving saccade to pursuit using image motion. *Int. Journal of Computer Vision*, 16(1):205–228, 1995.
- [44] D. Murray, I. Read, P. McLaughlin, and P. Sharkey. Reactions to peripheral image motion using a head/eye platform. In *Proc. of the Forth Int. Conference on Computer Vision*, pages 403–411, Berlin, Germany, May 1993. IEEE Computer Society Press.
- [45] K. Pahlavan. *Active Robot Vision and Primary Ocular Processes*. PhD thesis, Royal Institute of Technology, Stockholm, Sweden, May 1993.
- [46] K. Pahlavan, T. Uhlin, and J. Eklundh. Integrating primary ocular processes. In *Proc. of 2nd European Conference on Computer Vision*, pages 526–541, Santa Margherita Ligure, Italy, May 1992. Springer-Verlag.
- [47] K. Pahlavan, T. Uhlin, and J. Eklundh. Dynamic fixation. In *Proc. of the Forth Int. Conference on Computer Vision*, pages 412–419, Berlin, Germany, May 1993. IEEE Computer Society Press.
- [48] K. Pahlavan, T. Uhlin, and J. Eklundh. Dynamic fixation and active perception. *Int. Journal of Computer Vision*, 17(2):113–135, February 1996.
- [49] F. Panerai, C. Capurro, and G. Sandini. Space variant vision for an active camera mount. In *Proc. SPIE AeroSense95*, Florida, USA, April 1995.
- [50] M. Pobuda and C. Erkelens. The relation between absolute disparity and ocular vergence. *Biological Cybernetics*, 68:221–228, 1993.
- [51] P. Questa. A software simulation environment of a non-linear sampling retina-like ccd sensor. Technical Report 2/92, DIST-University of Genova, 1992.
- [52] I. Reid and D. Murray. Active tracking of foveated feature clusters using affine structure. *Int. Journal of Computer Vision*, 18(1):41–60, 1996.

- [53] D. Robinson. The oculomotor control system: A review. *Proc. of the IEEE*, 56(6), June 1968.
- [54] J. Santos-Victor. *Visual Perception for Mobile Robots: From Percepts to Behaviours*. PhD thesis, Universidade Técnica de Lisboa, Lisboa, Portugal, Novembro 1994.
- [55] J. Santos-Victor, G. Sandini, F. Curotto, and S. Garibaldi. Divergent stereo in autonomous navigation : From bees to robots. *Int. Journal of Computer Vision*, 14(2):159–177, 1995.
- [56] J. Santos-Victor, F. van Trigt, and J. Sentieiro. Medusa - a stereo head for active vision. In *Proc. of the Int. Symposium on Intelligent Robotic Systems*, Grenoble, France, July 1994.
- [57] E. Schwartz. Spatial mapping in the primate sensory projection : Analytic structure and relevance to perception. *Biological Cybernetics*, 25:181–194, 1977.
- [58] J. Semmlow, G. Hung, J. Horng, and K. Ciuffreda. Disparity vergence eye movements exhibit preprogrammed motor control. *Vision Research*, 34(10):1335–1343, 1994.
- [59] N. Takeda, M. Watanabe, and K. Onoguchi. Moving obstacle detection using residual error of foe estimation. In *Proc. of the 1996 IEEE/RJS International Conference on Intelligent Robots and Systems*, pages 1642–1647, Osaka, Japan, November 1996. IEEE Computer Society Press.
- [60] M. Tistarelli and G. Sandini. On the advantages of polar and log-polar mapping for direct estimation of the time-to-impact from optical flow. *IEEE Trans. on PAMI*, 15(8):401–411, April 1993.
- [61] F. Trigt, J. Santos-Victor, and J. Sentieiro. Design and construction of a stereo head for active vision. Technical Report 07/93, ISR / Instituto Superior Técnico, 1993.
- [62] T. Uhlin, T. Nordlund, A. Maki, and O. Eklundh. Towards an active visual observer. In *Proc. of the Fifth Int. Conference on Computer Vision*, pages 679–686, Massachusetts, U.S.A, June 1995. IEEE Computer Society Press.

- [63] R. Wallace, P. Ong, B. Bederson, and E. Schwartz. Space variant image processing. *Int. Journal of Computer Vision*, 13(1):71–90, September 1995.
- [64] J. Wang and E. Adelson. Layered representation for motion analysis. In *Proc. of the IEEE Computer Vision and Pattern Recognition Conference*, pages 361–366, New York, June 1993. IEEE Computer Society Press.
- [65] C. Weiman. Log-polar vision for mobile robot navigation. In *Proc. of Electronic Imaging Conference*, pages 382–385, Boston, USA, November 1990.
- [66] C. Weiman. Log-polar binocular vision system. Technical report, NASA, 1994.
- [67] J. Weng. Image matching using the windowed fourier phase. *Int. Journal of Computer Vision*, 11(3):211–236, December 1993.
- [68] F. Werblin, A. Jacobs, and J. Teeters. The computational eye. *IEEE Spectrum*, 1:21–29, May 1996.